Waseda University Doctoral Dissertation

Study on Multi-Scans based Invariant Feature Descriptors and Their Applications

Wei ZHOU

Graduate School of Information, Production and Systems
Waseda University

February 2013

Abstract

Pattern recognition systems, which are used to assist humans to replace human labor or solve problems that humans are ill-suited to address, have been increasingly blended into our daily life. Especially, face recognition with the property of non-intrusive, non-contact, high universality, high collectability, high acceptability, and low circumvention biometric, has been widely applied by our human beings, such as personal authentication, forensics, access control, security and so on. The academic study and commercial developments have kept increasing the rate of recognition precision and reducing the time complexity while decreasing the costs of equipment and hardwares. Although in literature pattern recognition systems have been shown powerful improvement, they are still considered to be unsolved due to the variations in scale, rotation, illumination, outliers, occlusion and so on.

In order to deal with the aforementioned challenges in the real-world recognition and classification problems, in this thesis, a new multi-scans encoding model and a new feature projection strategy are proposed. The multi-scans encoding model calculates the edge responses in multi-resolution and multi-orientation at every pixel location in the image and encodes a feature vector by the corresponding responses. A scale, rotation and illumination invariant feature descriptor can be obtained by using the direction of the most prominent edge response in related scale space. And our new feature projection strategy can characterize the within-class compactness and between-class separability under Fisher's linear discriminant function by L1-norm

and Correntropy optimization technique when outliers or occlusion occur.

In the multi-scans encoding model, multi-scans are applied to capture different spatial information on the image with advantage of less computation. The encoded feature descriptor is invariant to scale, rotation, illumination and more robust to outside factors change compared to traditional approaches. In order to capture global information of image and further reduce the effect of outside variations in pattern recognition, especially, facial expression and pose changes in face recognition, the multi-scans encoding model is extended into Curvelet frequency space. Our proposed Curvelet based invariant feature descriptors can overcome the weakness of traditional Gabor wavelets based ones in high dimensional space, and effective and efficient obtain the hyperplane and curved edge singularities, as a result, it can generate a feature descriptor with the property of high degree of anisotropy and orientation, high frequency resolution, which is suitable in lots of images with rich textures, such as facial images. The Curvelet and multi-scans encoding based feature descriptors are mutually complementary since Curvelet filter can capture the global information while multi-scans encoding based feature descriptors focus on the details of local appearance.

The new proposed feature projection strategy can solve the outliers and occlusion problems as well as select more compact and discriminative features. Two novel approaches called L1-norm based Linear Discriminant Analysis (LDA-L1) and Maximum Correntropy Criterion based Linear Discriminant Analysis (LDA-MCC) are proposed under this strategy. In general, LDA-L1 has several advantages: it is robust to outliers and occlusion as well as it is invariant to rotations. It also has no Small Sample Size problem, which usually exists in traditional feature projection methods. While LDA-MCC is capable of decreasing the influence of outliers greatly, especially, in the case of large outliers, resulting in a robust and powerful classification and it can be effectively solved by half-quadratic optimization algorithm.

Evaluations have been done on several large and famous face and texture databases with the applications of face identification, gender estimation, facial expression recognition and texture classification. The experimental results prove that the proposed approaches are superior to the state-of-the-art methods with the accuracy as well as speed.

Acknowledgements

I can not cover my happiness and proud in this moment. This is a key and amazing step in my life, especially in the research domain. I will share this achievement with all the people who assist, support or encourage me, since I can not get this award without them.

I would like to express my sincere thanks to all those who gave me the chance to finish this dissertation. Especially, I am deeply indebted to my supervisor Prof. Sei-ichiro Kamata, who gave me the chance to achieve my dream about doing the research in image processing and pattern recognition area. From the requirement to final implementation of the research and projects, he has given many awesome and wonderful ideas and advices to me, which not only how to solve the issue, but also why this issue comes. His effective and efficient encouragement and suggestions assist me all the moment during the period of researching and writing of this dissertation.

Besides my deal supervisor, my sincere thanks are also should be given to Prof. Takayuki Furuzuki, Prof. Takafumi Matsumaru, and Prof. Rin-ichiro Taniguchi, for their encouragement, insightful comments, and wonderful suggestions.

During these years, lab members gave me many supports. I want to gratitude them for all their support, help, interest and valuable hints. Furthermore, I will express my sincere pleasure to ones who I met in Japan, especially my girlfriend Pengyi Hao. Thanks to all of you accompanying with me in these wonderful and colorful years.

Last but most important, my special and deeply thanks is leading to my loving parents who never forget me and always give me powerful encouragement and support.

Contents

List of Tables

List of Figures

List of Symbols

Symbols

The global defined symbols are listed in the following.

D	Downsample factor
U	Upsample factor
θ	Direction of the scan
h	The interval of sampling points in the sequence
R	Radius of inscribed circle for one block
P	The number of sampling points in the block
$P^{'}$	The number of sampling points in the sequence
$P^{'}_{G}$	The number of sampling points in one group
L	Length of the sequence
N	The types of scans
S	The set of scan types
K	The number of learned codeword
S_b	The scatter matrix of between-class
S_w	The scatter matrix of within-class
Z_i	Position of the pixel in the block
$\widehat{Z}_i$	Position of the pixel in the sequence
A^T	Transpose of matrix A
A^{-1}	Inverse of matrix A
W	Projection matrix obtained by feature projection method
$\|\|.\|\|_{L1}$	L1 Norm
$\|\|.\|\|_{L2}$	L2 Norm
$\|\|.\|\|_{R1}$	R1 Norm

$L_a(.)$	LCBPMS map in layer a
$H(.)$	Histogram
$I(.)$	Intensity function
$Tr(.)$	Trace
$\mathcal{L}(.)$	Lagrangian function in construction of optimization problems
$g(.)$	Gaussian kernel function
$f_B(.,.)$	Binary encoding function
$f_T(.,.)$	Ternary encoding function
$f_Q(.,.)$	Quaternion encoding function
$k(x^{'}, y^{'})$	Kernel f
$p_l(.), q_l(.)$	Polarity functions
$pivot$	The center point in the sequence

1

Introduction

1.1 Research Background

Over the last several decades, pattern recognition has attracted substantial attention both in academia and industry. Many systems are applied in literature, such as face recognition, texture classification and so on. Thus, how to design such a system which is robust to the scale, rotation, illumination, outliers, occlusion and other variations has become the key issue in these developments and how to obtain robust, efficient and invariant features is the most important one. A general flow chart of patterns recognition, such as face recognition, texture classification, can be summarized in Fig.1.1. There are two key steps in this process, the first step is feature extraction. Some measurements of the image are calculated in this step, which form a feature descriptor. The image is then matched based on its feature descriptor and the dictionary which is obtained in the training process. Here, the definition of feature can be illustrated as the key characteristic of the samples or objects in an image. In general, the feature extraction should be performed in both learning and testing stages, and the features extracted from the testing set should be compared

with the ones from the learning set for further analysis.

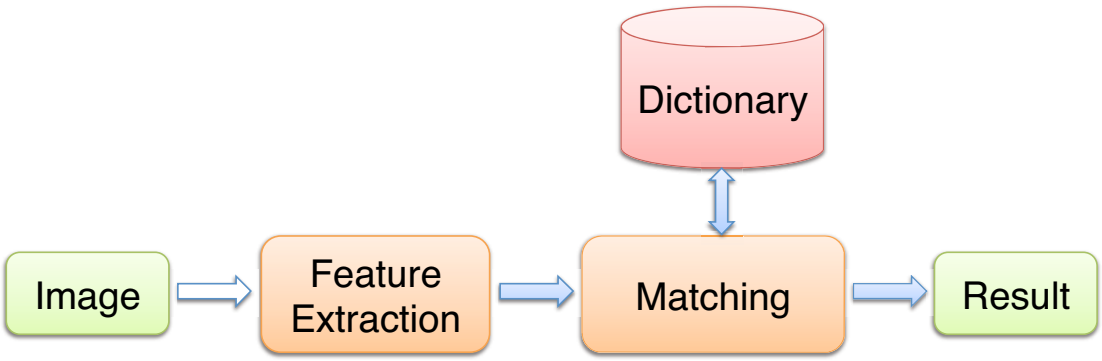


Fig. 1.1: A general flow chart of pattern recognition.

This thesis focuses on developing some novel invariant feature descriptors for effective and efficient pattern recognition, especially face recognition and texture classification.

1.1.1 Feature Descriptors

In the recent literature, many feature descriptors have been studied. In ref. [3], Scale Invariant Feature Transform (SIFT) was proposed to extract effective and discriminative local image features that are not only robust to image rotation and scale, but also partially invariant to the differences in illumination or viewpoint. In order to speed up the time computation, Bay et al. [4] proposed Speeded Up Robust Features (SURF) which could get a box filter approximation of second-order Gaussian partial derivatives based on integral images that just need near constant time to compute the rectangular box filters. On the other hand, the Gradient Location Orientation Histogram (GLOH) [5] was proposed to be an extension of SIFT that computes the histogram using a log-polar spatial grid and reduces the descriptor size by Principle Component Analysis (PCA). The main objective of GLOH is to improve matching results by using a more robust spatial grid to compute the gradient histogram. In ref. [6], Local Binary Patterns (LBP) was proposed by Ahonen et al. It

encodes a block of each pixel in an image as a pre-defined pattern, which is encoded as a binary number that thresholds the neighborhood values of every pixel which are compared with the center one. LBP is very sensitive to rotation change and in order to take rotation invariant into account, Liao et al. [7] proposed Advanced Local Binary Patterns (ALBP). The basic idea of this approach is performing a anti-clockwise circular shift for the bit number bit by bit and selecting the smallest decimal number. In ref. [8], Vu et al. proposed Patterns of Oriented Edge Magnitudes (POEM), which applies the LBP based structure into oriented edge responses.

Recently, since Gabor wavelet shows good characteristics in space frequency, space position, direction selectivity and captures global information of image effectively, feature descriptors based local patterns combined with Gabor feature have also been proposed to represent the image, such as Local Derivative Pattern(LDP) [9], Local Gabor Binary Patterns (LGBP) [10], Histogram of Gabor Phase Patterns (HGPP) [11] and Learned Local Gabor Patterns (LLGP) [12]. However, the common problem of these methods is that the feature dimension is very large due to Gabor decomposition, and Gabor transform cannot well represent curve singularity of human facial images since Gabor wavelets are very powerful to represent objects with isolated point singularities, but they are failed to present line or curve singularities.

In order to select more discriminative features as well as reduce the feature dimension, the subspace projection methods are widely combined withe the previous feature descriptors. PCA-SIFT [13] applied PCA [14] into SIFT features while FLD-SIFT [15] combined the Linear Discriminant Analysis (LDA) [16] with SIFT features. In ref. [17], Chan et al. projected high-dimensional LBP features into a discriminative space by LDA, and Joshi et al. [18] selected LBP features by Direct LDA. Shan et al. [19] combined piecewise LDA with LGBP features, the image is firstly encoded into LGBP map and then the entire LGBP feature vector is divided into segments

and LDA is applied into every segment separately.

The above mentioned subspace projection methods can reduce the feature dimension efficiently, however, all of them utilize L2-norm to measure the distances, which is sensitive to the presence of outliers or occlusion. Thus, the process of training maybe dominated by outliers or occlusion since the measurement is computed by summation of squared distances. In order to solve the outliers or occlusion issues, Dreuw et al. [20] used RANSAC into SIFT and SURF feature descriptors to partially remove the outliers' effect. In feature projection methods, L1 norm is introduced to PCA to eliminate the influence of the outliers or occlusion. And two novel approaches called L1 norm based PCA (PCA-L1) [21] and L1 norm based two-dimensional PCA (2DPCA-L1) [22] were proposed. Another state-of-the-art subspace approach called Linear Discriminant Analysis using Rotational invariant L1 norm (LDA-R1) [23] characterized the inter-class separability and the intra-class compactness by the rotational invariant L1 norm. Although these methods can improve the accuracy of pattern recognition or classification under small outliers or occlusion situation, the performance will decreased significantly when large outliers or occlusion occur.

1.1.2 Face Recognition

Followed by the improvement of our daily life, security has been taken more and more attention. Among many technologies, face recognition becomes one of the most key and effective biometric identification. For example, compared to fingerprint or iris, it is non-contact and non-invasive. In addition, face can also be easily recognized in the crowd. Since from the decision of our human beings, the first and important impression of people is based on someone's face [24]. What's more, by judging face, not only who you are can be got, but also more useful information can be obtained, such as the gender, facial expression and so on. As published in Ref. [25], compared to the other biometrics, such as voice, fingerprint, hand, eye and signature, face ranks

the first place in the principle of enrollment, redundancy, renewal, public perception, performance, storage in Machine Readable Travel Document (MRTD) system. Due to a lot of advantages of face recognition, it becomes more and more powerful in academia and industry. For instance, in many top-level conferences and journals, which are related to image processing, pattern recognition or computer vision; face recognition is one of the hottest topics. And more, many useful applications and increasing requests, such as arresting the criminal in public based on face, facial identification system, have been attracted more researchers and companies to develop novel approaches.

1.1.3 Texture Classification

Texture is one of the most important properties for an image that is usually defined as the local statistical property of a region with variable or constant patterns. Texture classification is defined as assigning an unknown texture image to one of a set with known texture labels. Therefore, it is a supervised classification where the output is a known class. Up to now, there are many applications related to texture classification both in academia and industry, such as pattern recognition, computer vision, remote sensing and so on. Texture classification has become the key to solve these public issues.

In general, the procedure of texture classification consists of two stages: learning stage and testing/recognition stage. Some key features or models should be extracted or constructed from the gallery with known class labels in the learning stage, which can form a database. The most useful low-level features are contrast; orientation and histogram while the model can be a discriminative function or probability curve. In the testing/recognition step, first, the features of the unknown sample are analyzed, and then these features are judged according to the database and passed through some classification algorithm to decide which category it should be assigned.

1.2 Motivation

In the literature, many researchers have proposed various pattern recognition approaches and systems, such as face recognition and texture classification, which are widely used in public to assist and benefit human beings. However, all these systems have been performed well under the controlled environment, but cannot reach a satisfactory level performance in the un-controlled condition. For example, all these applications face the challenges with the variations in scale, rotation, illumination, outliers, occlusion etc. A more reliable system should be deal with all these issues to obtain more reasonable and better performance. Among these issues, the major challenges can be summarized as follows:

1.2.1 Scale and Rotation

In pattern recognition, scale and rotation are two important aspects to evaluate the final result in real applications. For example, images can be obtained from different distances and orientations, but the same pattens can be presented in whatever distance or direction they are shot. During our daily life, it is very popular to resize or rotate the images for effective and efficient sharing and viewing. Human beings can deal with this problems easily but they are more difficult for computers.

1.2.2 Illumination

Firstly, different illuminations can generate various colors, for example, in the low color temperature, the white color patch becomes warmer while in the high color temperature, the same white patch is much cooler. That is, the intensity of the same object can be different under various illuminations. Secondly, due to the 3-dimension shape, the different direction of the same illumination can get various shadows and shadings, which will reduce the precision of feature representation. Thirdly but not the last, over-exposure or under-exposure will change the texture information in the

surface of the object.

1.2.3 Outliers

Outliers can be defined as an observation that is numerically distant from the rest of the data, which lie outside the overall pattern of a distribution. For example, the data may have been added incorrectly. This can be a case which does not fit the model under study, or an error in the measurement. Usually, the presence of outliers indicates some sort of problem, such as inaccurate data fitting, false recognition. Thus, in a robust pattern recognition system, it is significant to remove the influence of outliers.

1.2.4 Occlusion

In our daily life, it is common to see that one object is covered by another one. For example, in the crowd, our face is easily occluded by the other face or body. For our human beings, we can recognize an object even it is partially occluded by other object, since our brain can easily separate these occluded objects by experience. However, it is very difficult to deal by computers. Since in this scenario, the extracted features will be wrong if the occluded part is included, while the features will be decreased power if only partial true object is used. Thus, no matter which kind of case, the final recognition rate will be reduced.

1.2.5 Speed

In order to apply recognition technology into our real day life, we should build a real-time recognition system. However, most features obtained by pattern recognition approaches are located in a very high dimension, such as Gabor based one. The time complexity is so high that these features can not be applied in our real-time applications. Thus, how to design or extract fast and reliable features is a key issue. In general, this kind of high dimensional data contains high redundant information

and the intrinsic structure of the data maybe lied into a low dimension space. This property leads researchers to develop some dimension reduction methods that can extract the discriminative information from original high dimensional space.

1.3 Principle Contributions

This work is targeted towards the development of the invariant and discriminative feature descriptors. In this thesis, in order to tackle the aforementioned challenges, we propose a new multi-scans encoding model and a new feature projection strategy. The block extracted from a image will be scanned by several basic scan models which are defined as Fig.1.2. Here, the black rectangle stands for the block, the blue line is the scan order, the purple circle is the boundary for the scan and parameters R and θ stands for the size of this block and direction of the scan, respectively. The starting point of the basic scan model is on the position $(0, -R)$ and ending point is $(0, R)$ with the starting direction of θ. Multi θ values can get multi-orientation of this block, where $\theta = [\theta_1, \theta_2, ..., \theta_N]$, $\theta_n = \Delta\theta \times (n-1)$ and $\Delta\theta = \frac{360^o}{N}(n = 1, ..., N)$. Based on value of θ_n, N types of scans can be obtained. And with different θ_n values on the basic scan model, the original 2D block can be transformed into varying 1D sequences with the length of L, which capture different direction information of this block. For a fixed θ_n, upsampling and downsampling will be operated in the related 1D sequence to get multi-resolution representation of the sequence like Fig.1.3. Here, the sequence is upsampled by a factor U and downsampled by a factor of D. I is the intensity, $Idx_j (j \in \mathbb{N})$ stands for the index of sampling point in the related resolution sequence. If $j < 0$, then the related sequence is downsampled, while $j > 0$ means the related sequence is upsampled. And Idx_j is defined as

$$Idx_j = \begin{cases} 1, 2, ..., \lfloor (L-1)/|j| \rfloor + 1 & \text{if } j < 0, \\ 1, 2, ..., L & \text{if } j = 0, \\ 1, 2, ..., j(L-1) + 1 & \text{if } j > 0. \end{cases} \tag{1.1}$$

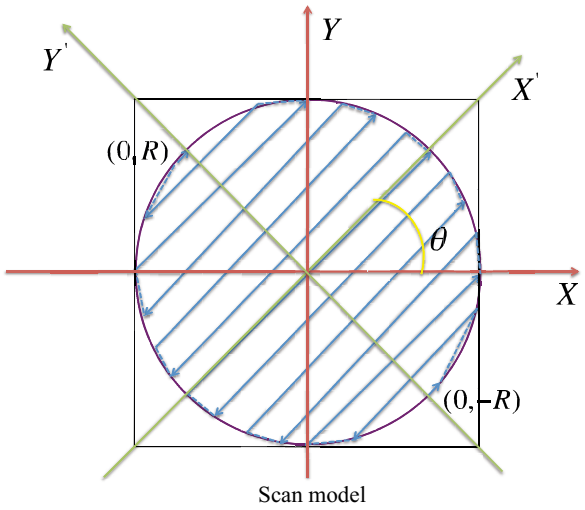


Fig. 1.2: The basic scan model.

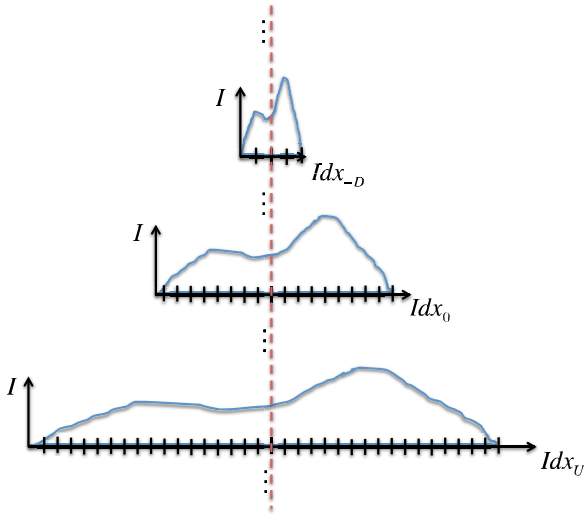


Fig. 1.3: Multi-resolution of one block scanned with same orientation by downsampling and upsampling.

One thing should be pointed out is that the center sampling point in all sequences for one block is same. And the neighborhood points around the center point is upsampled and downsampled. After obtaining the multi-resolution representation of the original 2D block shape with a fixed θ_n, we will calculate the edge response of them. The sequence which has strongest edge response is selected to encode, resulting in a code number. By considering all the orientations of one block and all the blocks in the image, at last, we can obtain a histogram as the feature descriptor of this image where horizontal coordinate is the index of code and vertical coordinate is the frequency of the code.

Our new feature projection strategy under Fisher's linear discriminant function by L1-norm and Correntropy optimization technique can deal with the observation (outliers or occlusion) that does not obey the pattern of the majority of the data. This kind of observation is considered as the error estimate of Fisher's linear discriminant function as shown in Fig.1.4(a), where blue diamonds stand for data and yellow triangles are outliers. L1-norm and Correntropy variance optimization can make these error have little or no influence to the Fisher's linear discriminant function. For example, in Fig.1.4(a) that is related to the original feature space, the feature distribution of data and outliers are merged. After projecting them into projected feature space by our proposed feature projection strategy, the outliers and data can be separated much easily by the purple line in Fig.1.4(b).

In a word, the multi-scans encoding model computes the edge responses in multi-resolution and multi-orientation at every pixel location in the image and encodes a feature vector by the corresponding responses. A scale, rotation and illumination invariant feature can be obtained by using the direction of the most prominent edge response in related scale space. And our new feature projection strategy can characterize the within-class compactness and between-class separability under Fisher's linear discriminant function by L1-norm and Correntropy optimization technique

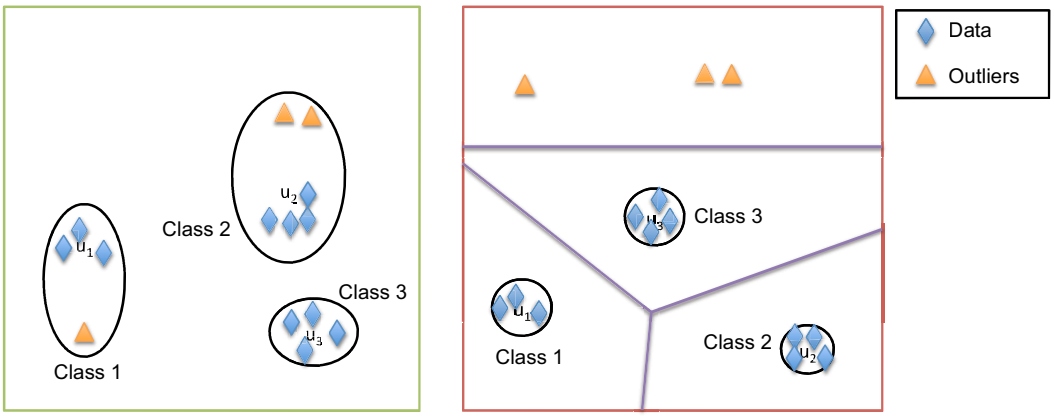


Fig. 1.4: Blue diamonds stand for data and yellow triangles are outliers. (a) The feature distribution of three classes with outliers in original feature space. (b) The feature distribution of three classes and outliers in the projected feature space by proposed feature projection method.

when outliers or occlusion occur.

The comparison of our proposed methods and some famous approaches mentioned above are listed in Table 1.1. Here, ∘ means that a method is invariant to the related variation, ◊ means that a method is a little sensitive to the related variation while × means that an approach is very sensitive to the related variation. Note that here, in"Li et al." and"Pang et al." approaches, the feature descriptors are based on Gabor space instead of traditional intensity based one. From this table, we can recognize that the method by"Pang et al." in the state-of-the-art approaches can deal with all the problems. However, two issues should be pointed out for this approach. The first one is that high dimension Gabor feature used in this method failed to represent line or curve information effectively. The second one is that 2DPCA-L1 used in this approach is good at reconstruction problem, but is not well performed in obtaining discriminative features for further classification. And our proposed method in chapter 3 will solve these two points with much better performance.

Table 1.1: Comparison of our proposed methods and some famous approaches.

Method \ Variation	Scale	Rotation	Illumination	Outliers	Occlusion
SIFT [3]	○	○	◊	×	×
SURF [4]	○	○	○	×	×
GLOH [5]	○	○	○	×	×
LBP [6]	○	×	◊	×	×
POEM [8]	○	×	○	×	×
Dreuw et al. [20]	○	○	○	◊	◊
LGBP [10]	○	○	○	×	×
Li et al. [23]	○	○	○	◊	◊
Pang et al. [22]	○	○	○	○	○
Our method in Chapter 2	○	○	○	×	×
Our method in Chapter 3	○	○	○	○	○

1.4 Organization

The remaining chapters are organized in the following.

Chapter 2 proposes some novel invariant feature descriptors for presenting the local features of digital image using 1D Local Patterns by Multi-Scans (1DLPMS), which are scale, rotation and illumination invariant. Based on the multi-scans model, the block extracted from image will be transformed into several sequences with multi-resolution and multi-orientation representation. Some specific sequences will be selected to encode based on their edge responses. In addition, the simplifications of our proposed approaches into images analysis are also considered. Our proposed approaches are expected to be less sensitive to outside changes besides preserving the appearance of images embedded in the original gray scale. Simple evaluation based on one general database will test the performance of our proposed feature descriptors.

Chapter 3 proposes some Curvelet transform based invariant feature descriptors for presenting the local features of images, which are called Local Curvelet Binary Patterns by Multi-Scans(LCBPMS) and Learned Local Curvelet Patterns (LLCP). The feature descriptors in this chapter can be seen as the Curvelet frequency ex-

tension of feature descriptors in chapter 2. Our proposed methods can overcome the weakness of traditional Gabor wavelets in higher dimensions, and capture the curve singularities and hyper plane singularities of images, such as facial images. The Curvelet and multi-scans based feature descriptor are mutually complementary because multi-scans based feature descriptor captures the local appearance detail, whereas Curvelet extracts global information. In order to select more compact and discriminative features and solve outliers or occlusion problem as well as reducing the dimension of our proposed feature descriptor, one novel feature projection method named L1-norm based Linear Discriminant Analysis (LDA-L1) is proposed. Our proposed feature projection approach is capable of decreasing the influence of outliers or occlusion significantly, giving a powerful and robust classification. At last, performance assessment under some variations will be evaluated between our proposed approaches and some traditional ones.

Chapter 4 studies face recognition problems by our proposed multi-scans based invariant feature descriptors. By considering that the face is expressed in 2D space, the feature projection method in chapter 3, which is based on a vector representation, is firstly extended into matrix based one to capture the spatial structure information in our face. And then one effective classifier called Weighted Histogram Spatially constrained Earth Mover's Distance (WHSEMD), which illustrates the discriminative powers of various image regions, the different patterns and the different spatial information of images, is proposed. In the final, huge experiments on applications of face identification, gender estimation and facial expression recognition will be tested our proposed approaches compared to the conventional approaches in terms of both accuracy and speed.

Chapter 5 describes a general texture classification application using our proposed feature descriptors. In order to further solve large outliers problem, another novel method of Linear Discriminant Analysis (LDA) based on a new Maximum Cor-

rentropy Criterion (MCC) optimization technique is proposed, which can solve large outliers problem compared to other feature projection methods. Our new method called MCC based LDA (LDA-MCC) is capable of analyzing non-Gaussian noise to reduce the influence of large outliers substantially while keeping the most discriminative information, and it can be effectively solved by half-quadratic optimization algorithm. At last, experiments are evaluated based on two famous and challenging texture databases with the variations of scale, rotation, illumination and outliers.

Chapter 6 summarizes this dissertation and some future works are list.

2

Invariant 1D Local Feature Descriptors by Multi-Scans

2.1 Introduction

As mentioned in chapter 1, our multi-scans encoding model is targeted to extract the local feature of an image. The basic idea is that it assigns several sequences to each pixel in an image and considers the relationship between the center pixel(we call it *pivot* in the following) and neighborhood ones. These sequences are transformed from the block in the image by multi-scans and Fig.2.1 gives a generalized definition for the block of size $(2R+1) \times (2R+1)$. Here, $\alpha = (2R-1) \times (2R-1)$. First, let Z_0 be the center of the block. Then in each rectangle with inscribed circle of radius of R (the center of the circle is also Z_0), we can mark this rectangle from left-top as Z_α, $Z_{\alpha+1}$, ..., $Z_{\alpha+8R-1}$ by clockwise order.

In the literature, the most famous local feature descriptor called LBP (details see [6]) was proposed some years ago. Generally, the symbol (P, R) is applied to present LBP that denotes P equally sampling points on a rectangle with inscribed radius of R in circle. The advantage of LBP is that it is less sensitive to the monotonic

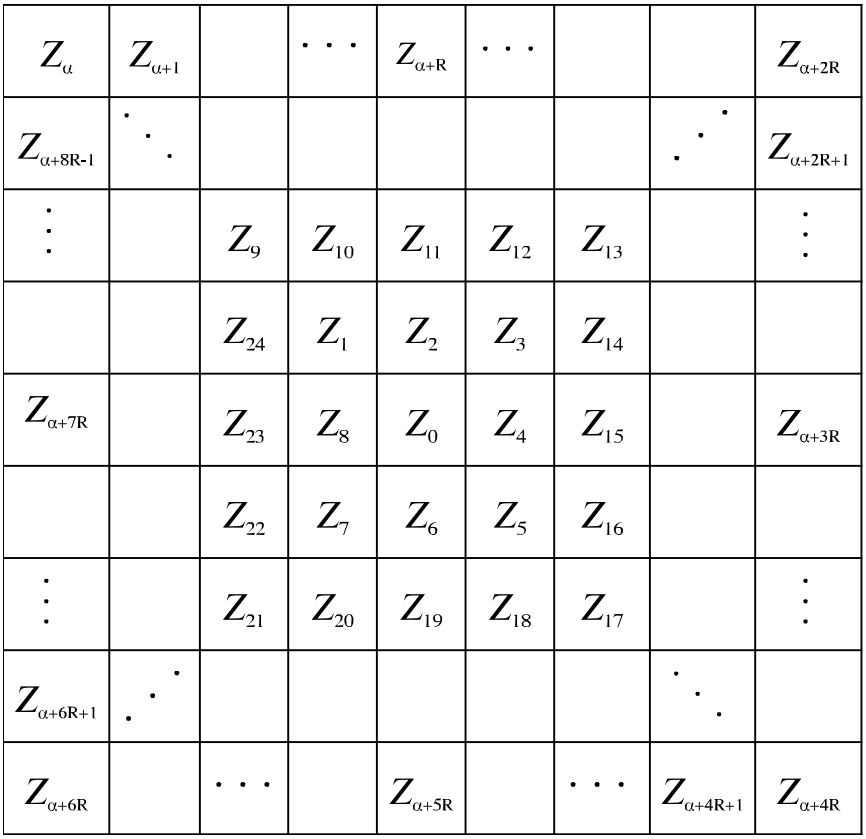


Fig. 2.1: Block of image pixel with the size of $(2R+1)$ x $(2R+1)$.

illumination changes, however, it always suffers more in non-monotonic illumination changes and random noise. In order to overcome this issue, Local Ternary Patterns(LTP) [26] and Local Quaternion Patters(LQP) [27] are studied in our work. The adaptive thresholding function $f_B(.,.)$, $f_T(.,.)$, $f_Q(.,.)$ called binary, ternary, quaternion encoding rule for the basic LBP, LTP and LQP can be formally represented as

$$f_B(I(Z_0), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) \leq I(Z_0), \\ 1, & \text{if } I(Z_i) > I(Z_0), \end{cases} \tag{2.1}$$

$$f_T(I(Z_0), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) < I(Z_0) - interval, \\ 1, & \text{if } I(Z_0) - interval \leq I(Z_i) \\ & \quad \leq I(Z_0) + interval, \\ 2, & \text{if } I(Z_i) > I(Z_0) + interval, \end{cases} \tag{2.2}$$

$$f_Q(I(Z_0), I(Z_i)) = \begin{cases} 0, & \text{if } I(Z_i) < I(Z_0) - interval, \\ 1, & \text{if } I(Z_0) - interval \\ & \quad \leq I(Z_i) \leq I(Z_0), \\ 2, & \text{if } I(Z_0) < I(Z_i) \\ & \quad \leq I(Z_0) + interval, \\ 3, & \text{if } I(Z_i) > I(Z_0) + interval, \end{cases} \tag{2.3}$$

where Z_i is the neighborhood point around $Z_0(pivot)$. $I(Z_i)$ and $I(Z_0)$ are the intensity of the neighborhood points and $pivot$.($I(Z_i), I(Z_0) \in \mathbb{R}$ and $0 \leq I(Z_i), I(Z_0) \leq \widehat{MI}$, $\widehat{MI}$ is the maximal intensity in the image) The definition of $interval$ is notated as Eq. (2.4). τ is a predefined value [26].

$$interval = \begin{cases} I(Z_0)/\tau, & \text{if } I(Z_0) \leq \widehat{MI}/2, \\ (\widehat{MI} - I(Z_0))/\tau, & \text{if } I(Z_0) > \widehat{MI}/2. \end{cases} \tag{2.4}$$

This kind of patterns can also be considered as the concatenation of the binary/ternary/quaternion gradient directions, and are called micro-patterns. The histograms of these micro-patterns encode the structure information of the edges, curves and other local distributions in a sample. Fig.2.2 and Fig.2.3 show some examples of obtaining LBP, LTP and LQP micro-patterns. In both figures, the upper 3 by 3 block is the pixel intensity from image and the lower ones are the corresponding local patterns. From Fig.2.2, we can see that, when some light source came from the bottom left corner, LBP micro-patterns are changed while LTP micro-patterns keep the same value. Thus, LTP is more insensitive to illumination compared to LBP. In Fig.2.3, it is clear that the left top 3 by 3 block stands for flat area while the right top 3 by 3 block stands for spot. However, LBP micro-patterns can not separate

them while LQP micro-patterns can discriminate them clearly. Thus, LQP can get more discriminative information about image compared to LBP.

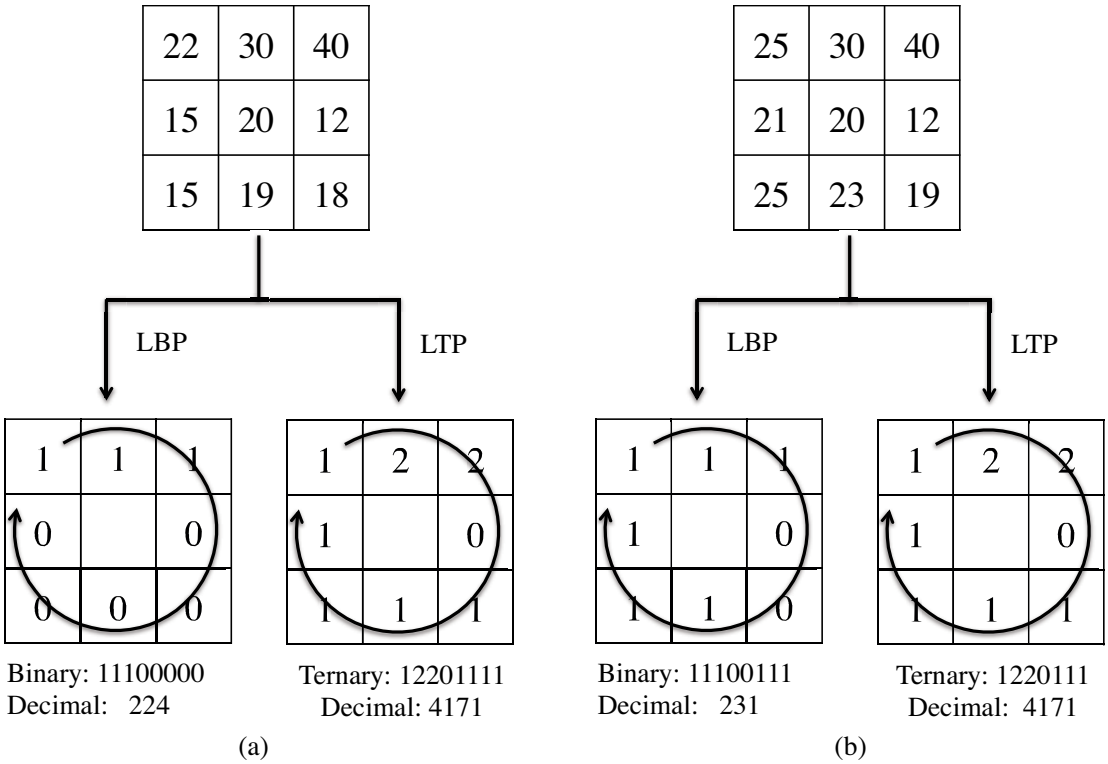


Fig. 2.2: LBP vs LTP, $\tau = 4$. (a) LBP and LTP operators, (b) LBP and LTP operators by the corresponding block in (a) under illumination change.

Some reconstructed images (here, facial images are used) by decimal number according to corresponding local patterns are illustrated in Fig. 2.4 (Note that all the decimal numbers are normalized into [0,255] for display). From these reconstructed images, we can clearly see that LTP and LQP are more invariant to illumination changes and can get the key parts in faces more efficiently than LBP, such as eyes, nose and mouth. However, one problem is that the number of micro-patterns obtained by LTP and LQP are larger than traditional LBP with the same bit number, which takes time for matching. This is one motivation for the following research. And another motivation is that this kind of local patterns are sensitive to rotation

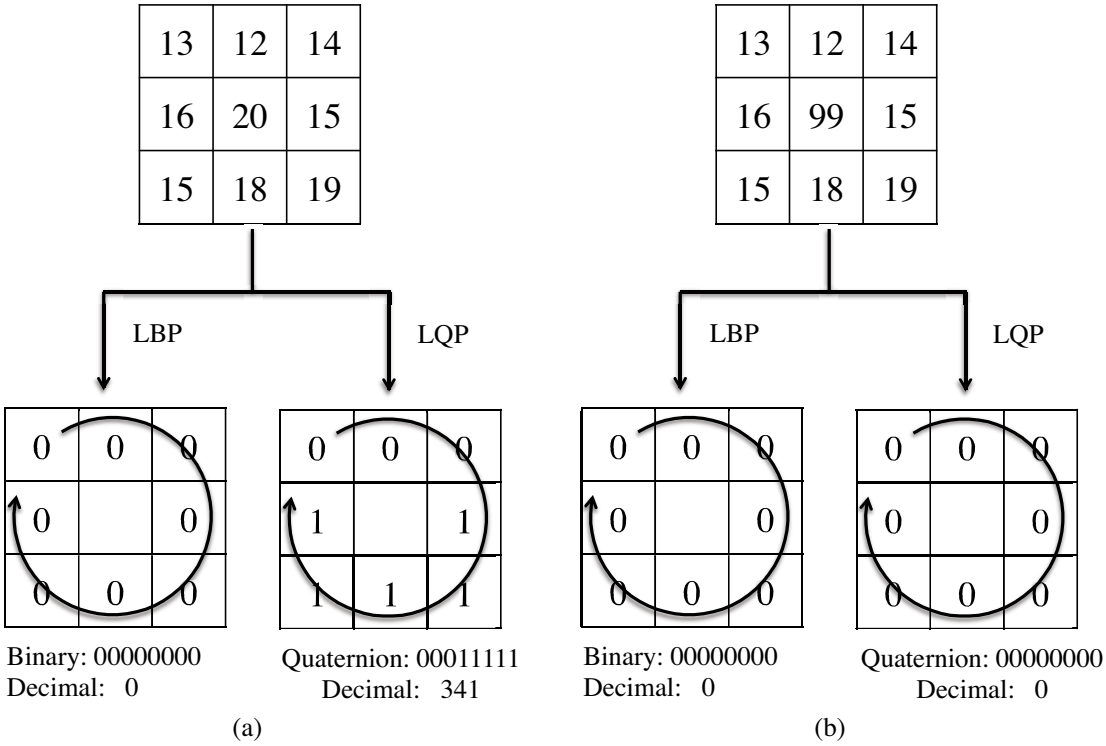


Fig. 2.3: LBP vs LQP, $\tau = 4$. (a) LBP and LQP operators for flat area, (b) LBP and LQP operators for spot.

change.

In order to take rotation invariant into account, Advanced Local Binary Patterns (ALBP) was proposed by Liao et al. [7]. The basic idea of this pattern is performing an anti-clockwise circular shift into the pattern numbers bit by bit and selecting the smallest decimal number. For example, in Fig.2.2(a), 8 binary numbers can be obtained and 00000111 is selected at last. However, this method could not solve the basic LBP problems with noise and illumination issues and has more time complexity than basic LBP approach. More recently, Dominant Local Binary Pattern(DLBP) [28] and Local Derivative Pattern(LDP) [9] were proposed. DLBP considers more complicated shapes, which could contain crossing boundaries, highly curved edges and corners. However, in our applications, such as face, most of shapes are fundamental, such as line, edge, spot and flat area. And the conventional LBP

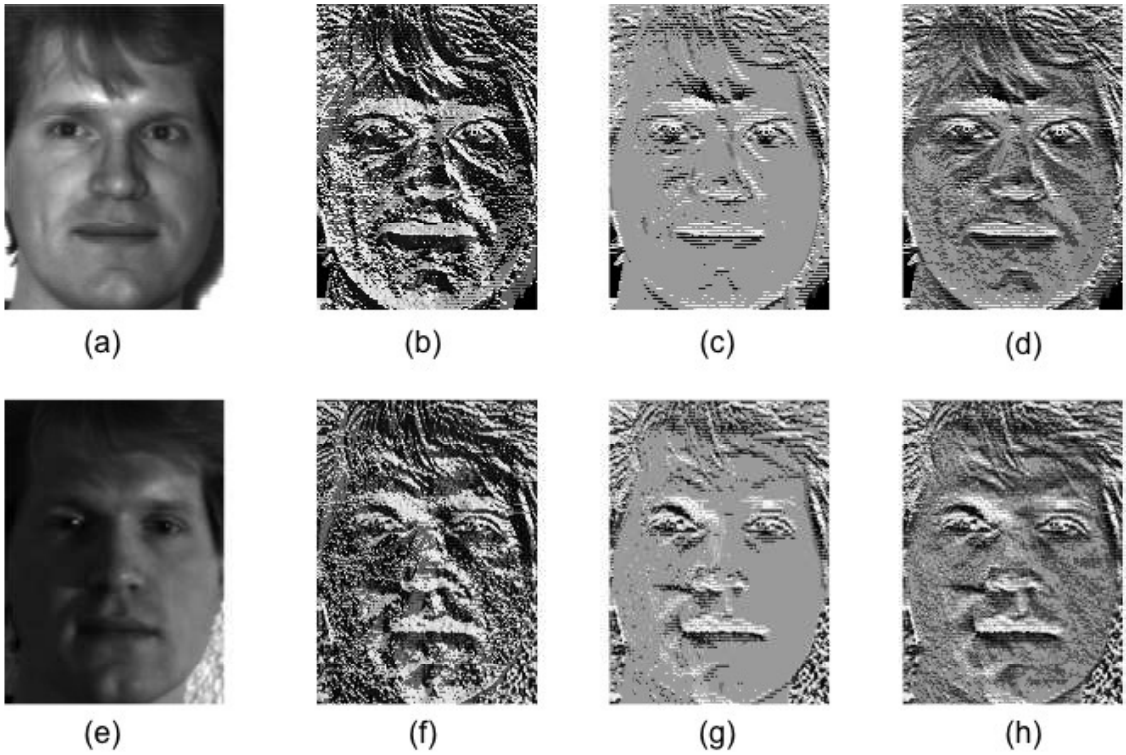


Fig. 2.4: Reconstructed images by LBP, LTP and LQP. (a), (e) are two original images under different illumination conditions. (b), (f) are two reconstructed images by LBP corresponding to (a) and (e), while (c), (g) are reconstructed images by LTP and (d), (h) are reconstructed images by LQP, respectively.

could catch about 90 percent of the shapes in case of preprocessed FERET facial images [6]. Thus, in our study, most fundamental information is considered since it is enough for face recognition or texture classification as well as it can obtain faster speed. LDP uses a higher order derivative structure of samples to construct the feature space; therefore, compared to LBP that only uses first order derivative structure of samples, it can capture more detailed and useful structure information. Thus, LDP contains more powerful and discriminative features than LBP. However, on the other side, LDP generated more patterns than LBP which make it difficult to be applied in practice. In some cases, higher-order derivative information also caught more noise information which is useless. Compared to these methods men-

tioned above, our proposed methods in the following, which are based on multi-scans encoding model, have amazing speed while keeping discriminative and fundamental information as well as they are invariant to scale, rotation, illumination and so on. The multi-scans encoding model calculates the edge responses in multi-resolution and multi-orientation at every pixel location in the image and extracts a feature vector from the corresponding strength responses. A scale, rotation and illumination invariant feature descriptor can be obtained by using the direction of the most prominent edge response in related scale space.

The remainder of this chapter is organized as follows: 1D Local Patterns by Multi-Scans will be introduced in section 2. Grouped 1D Local Patterns by Multi-Scans will be described in section 3 and evaluation will be estimated in section 4. Finally, conclusions are discussed in section 5.

2.2 1D Local Patterns by Multi-Scans (1DLPMS)

In this section, our proposed multi-scans encoding model is developed. Let S be a set of multi-scans. Based on Fig.1.2 and Fig.2.1, with different $\Delta\theta$ and R, we can generalize the orders of different scans. For example, Fig.2.5 for $\Delta\theta = 45^o$ and R=1 while Fig.2.6 for $\Delta\theta = 45^o$ and R=2 are two cases, the direction of the scans $\theta = [0^o, 45^o, 90^o, 135^o, 180^o, 225^o, 270^o, 315^o]$. In these cases, eight scans which are located in the circle with the radius of R are used. Thus, S={S1, S2, S3, S4, S5, S6, S7, S8}, the number of scan type $N = 8$. Based on the multi-scans, one block can be transformed into several sequences.

Since scan order itself contains the spatial information of the image while different scan orders can get different spatial information, only little neighborhood pixels encoding in multi-scans can also keep enough discriminative and spatial information in that block. For general, the symbol (P', h) will be applied for this neighborhood pattern that illustrates P' sampling points with a interval of h between two adjacent

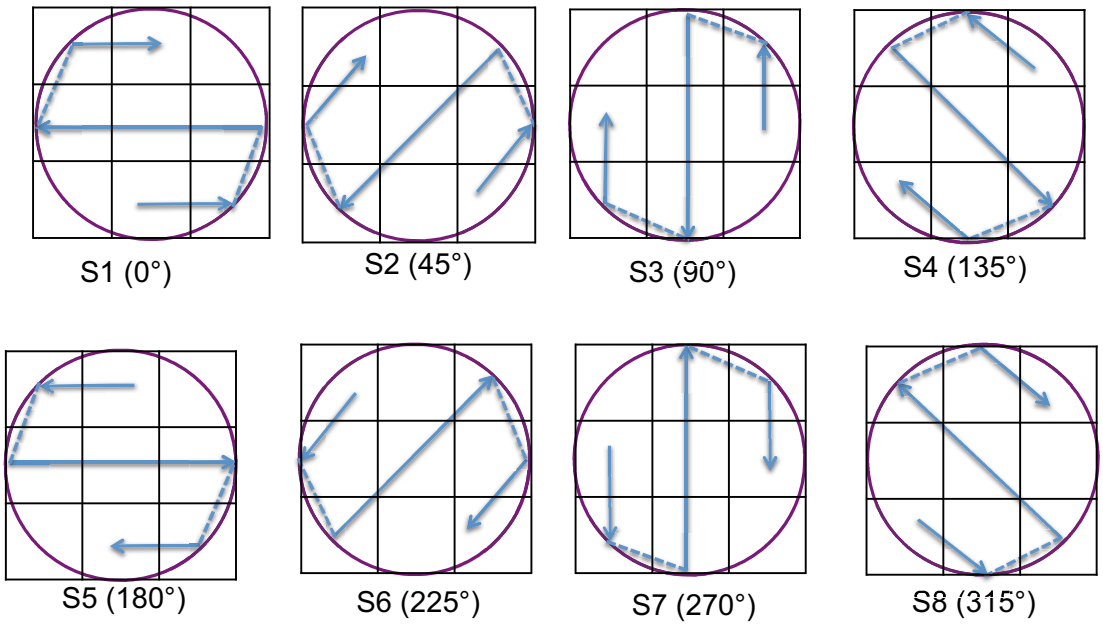


Fig. 2.5: Multi-scans for one block($\Delta\theta = 45^o$ and R=1).

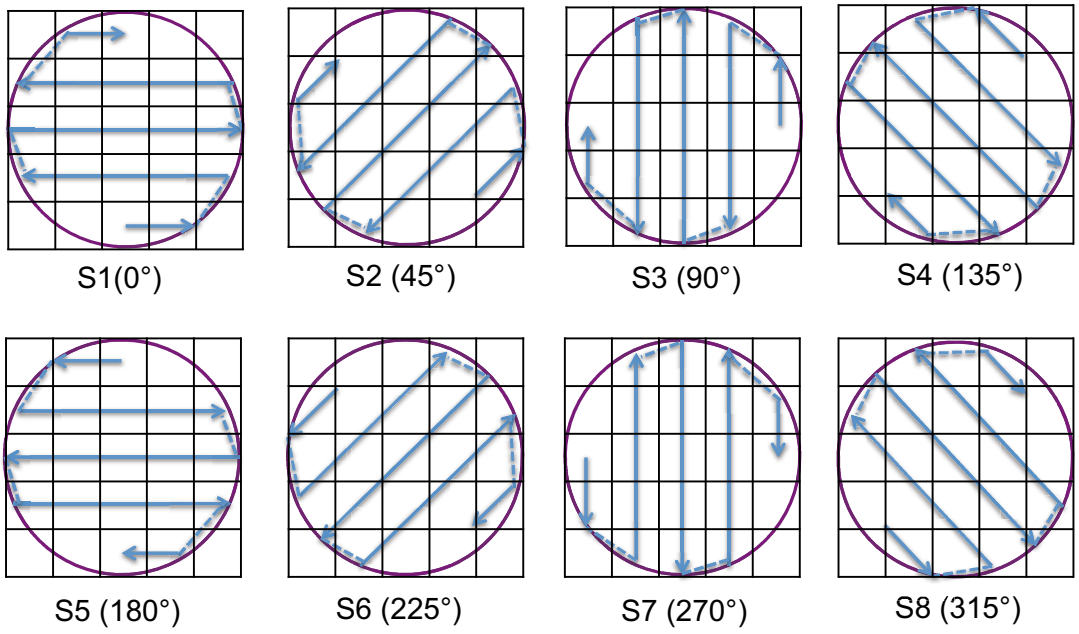


Fig. 2.6: Multi-scans for one block($\Delta\theta = 45^o$ and R=2).

sampling points.

Some traditional local patterns which are based on circle encoding, such as LBP [6], assign a circle to every pixel of an image for encoding to present the micro-patterns. However, when some distortion, such as illumination or noise, is imported in the block, the pattern of LBP will be changed as shown in Fig.2.7 where the dash pixel contains some distortion. Based on this point, multi-scans are applied for encoding in our study as illustrated in Fig.2.8 (four types of scans are used) where several patterns can be obtained. Compared Fig.2.8(a) with Fig.2.8(b), just one pattern is changed while other three patterns are same. In this scheme, encoding by multi-scans can not only keep the texture information but also reduce the effect of image distortion.

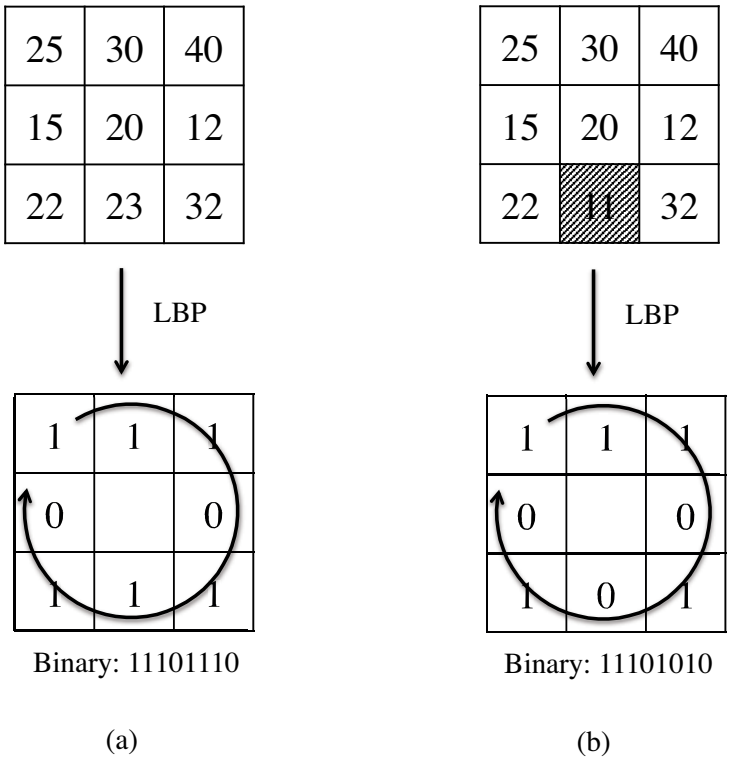


Fig. 2.7: Comparison of obtaining the LBP micro-patterns, the dash pixel is distorted.

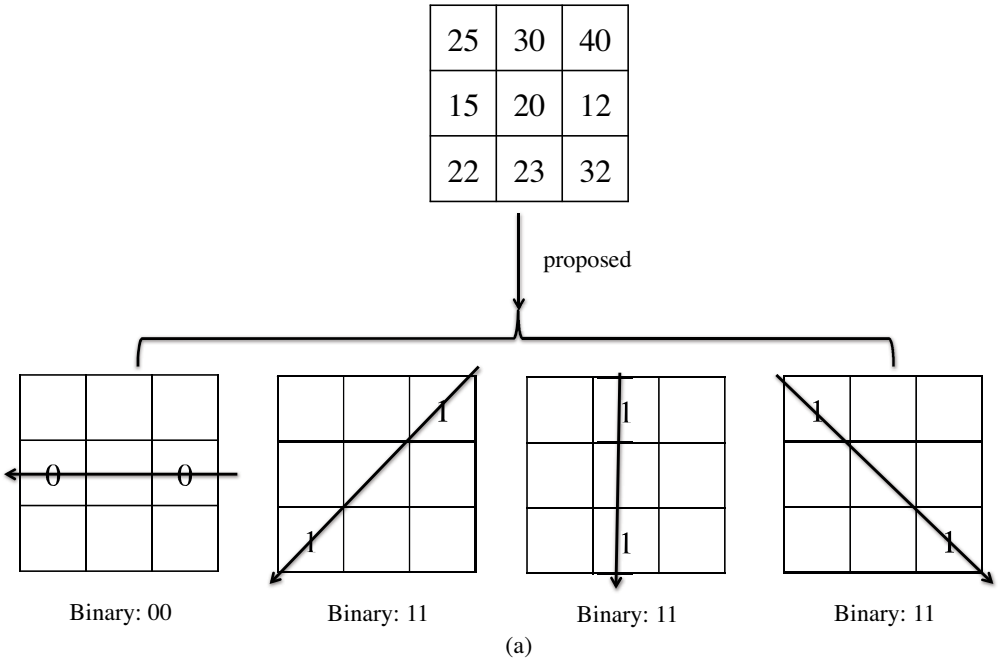


Fig. 2.8: Comparison of obtaining our proposed micro-patterns, the dash pixel is distorted.

On the other hand, just the circle spatial information around the pixel is considered in traditional methods. As shown in Fig.2.1, eight positions from Z_1 to Z_8

around Z_0 are taken into account and the number of obtained patterns is $2^8 = 256$ for $P = 8$ and $R = 1$. Generally speaking, the spatial information and the number of patterns are related to accuracy and time consuming respectively. In this study, when multi-scans are applied to encode the neighborhood pixels, more spatial information around Z_0 can be obtained while the number of patterns is reduced at the same time.

As mentioned above, there are eight locations marked as Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8 around Z_0 in LBP, while multi-scans are applied in this block($P^{'} = 4$, $h = 1$, $\Delta\theta = 45^o$ and $R = 2$), S1 can cover position Z_{23}, Z_8, Z_4 and Z_{15} , S2 can cover position Z_{21}, Z_7, Z_3 and Z_{13}, S3 can cover position Z_{19}, Z_6, Z_2 and Z_{11}, S4 can cover position Z_{17}, Z_5, Z_1 and Z_9, S5 can cover position Z_{15}, Z_4, Z_8 and Z_{23}, S6 can cover position Z_{13}, Z_3, Z_7 and Z_{21}, S7 can cover position Z_{11}, Z_2, Z_6 and Z_{19} while S8 can cover position Z_9, Z_1, Z_5 and Z_{17} (Z_0 is set at the center of each scan). In this case, the number of patterns is just $2^4 = 16$, which is much less than LBP which needs $2^8 = 256$ patterns. And multi-scans can cover 16 positions while circle can just cover 8 positions. Therefore, we can use less number of patterns to represent enough discriminative and more spatial information in that block. In other words, multi-scans can not only keep more spatial information than LBP, but also use less number of patterns. Note if $P^{'} = 6$, then S1 can cover position Z_{22}, Z_{23}, Z_8, Z_4 Z_{15}, Z_{14}, while S2 can cover Z_{20}, Z_{21}, Z_7, Z_3, Z_{13}. Z_{12} and so on.

In order to get multi-resolution representation of one sequence in related orientation, upsampling by a factor $U$and downsampling by a factor D are applied into the sequence. Fig.2.9(a) illustrates an example where one block of image is transformed into one sequence by S1 (for simplicity, the intensity of starting and ending points are 22 and 40) and Fig.2.9(b) is the intensity curve of the corresponding sequence with length L ($L = 9$). Fig.2.10 shows the multi-resolution representation of the sequence in Fig.2.9, where $D = 2$ and $U = 2$. The related intensity curve is drawn

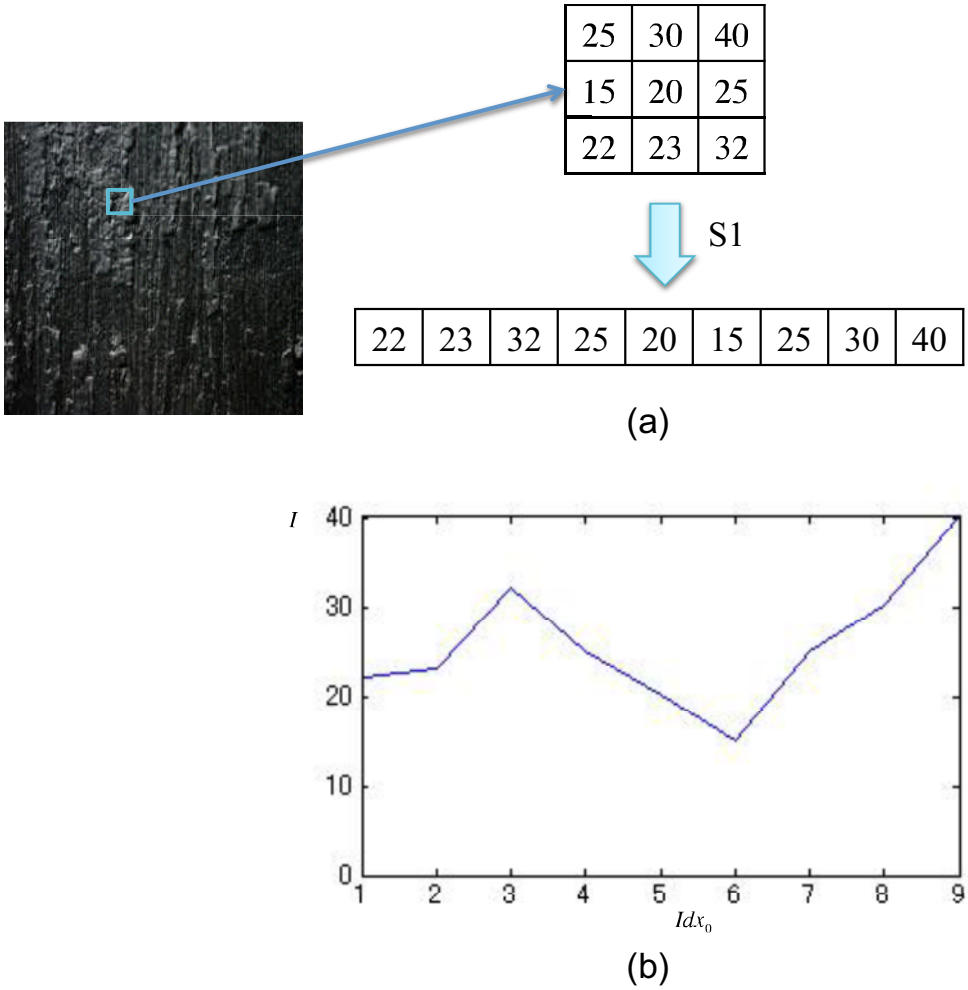


Fig. 2.9: Representation of a block by one type of scan. (a) Transform a block into one sequence by S1. (b) Intensity curve of the corresponding sequence.

near each sequence. Here, index of sequence in each resolution is $Idx_0 = [1, 2, ..., 9]$, $Idx_{-2} = [1, 2, 3, 4, 5]$ and $Idx_2 = [1, 2, ..., 17]$. After multi-resolution representation of the sequence in one orientation, we will select one which has the strongest edge response to make scale invariant property in that orientation. Firstly, for each position $\widehat{Z}_i$ in the sequence, absolute gradient is calculated as $\widehat{g}(\widehat{Z}_i) = |I(\widehat{Z}_{i+1}) - I(\widehat{Z}_{i-1})|$, let $L_D = \lfloor (L-1)/D \rfloor + 1$ be the length of the sequence when it is downsampled by a factor D and $L_U = U(L-1) + 1$ be the length of the sequence when it is

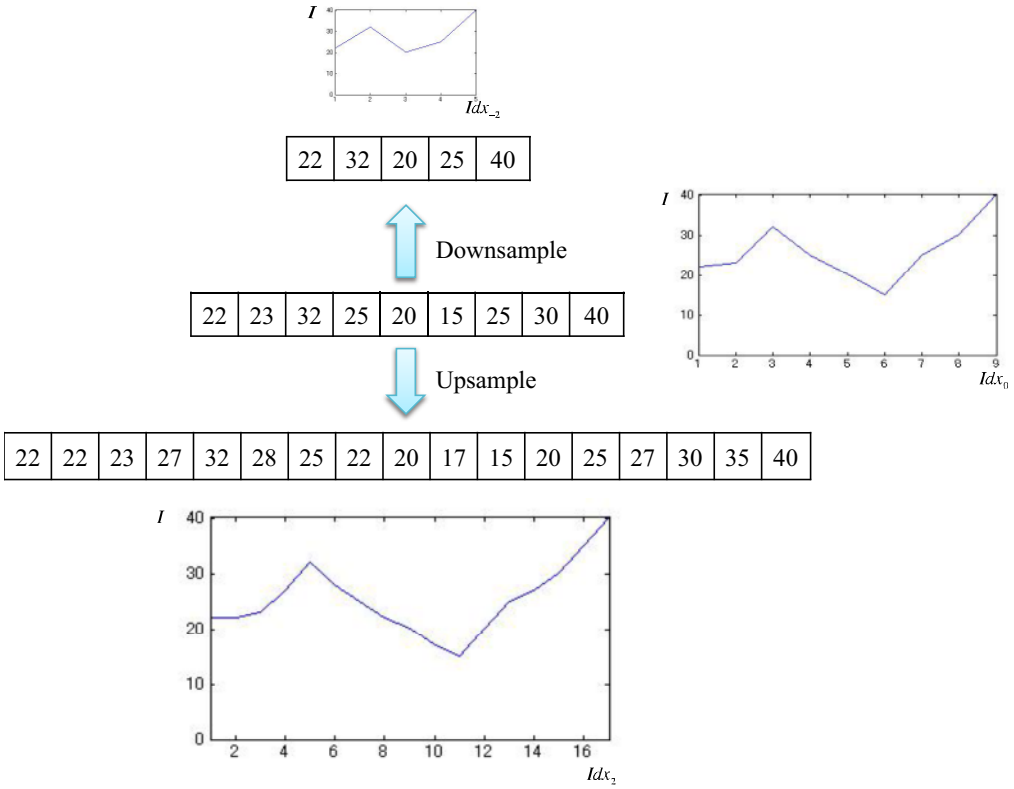


Fig. 2.10: Multi-resolution representation of one sequence.

upsampled by a factor U. Then the edge response of the multi-resolution sequences with orientation θ is computed as

$$ER^{\theta}(D,U) = \begin{cases} \dfrac{1}{L_D}\displaystyle\sum_{i=1}^{L_D}\widehat{g}(\widehat{Z}_i), & \text{if the sequence is downsampled by a factor } D, \\ \dfrac{1}{L_U}\displaystyle\sum_{i=1}^{L_U}\widehat{g}(\widehat{Z}_i), & \text{if the sequence is upsampled by a factor } U\ , \\ \dfrac{1}{L}\displaystyle\sum_{i=1}^{L}\widehat{g}(\widehat{Z}_i), & \text{others.} \end{cases} \tag{2.5}$$

Then, for each scan, we can get one sequence with the strongest edge response in the related resolution. When multi-scans are applied into one block, multi sequences

can be got. And the top $\xi - th$ strongest edge response sequences in all orientations will be selected for further encoding.

In general, in order to encode the selected sequence, several reference patterns are defined in advance. Some reference patterns for $P^{'} = 2$ and $h = 1$ are listed in Fig.2.11. Then, one block can be represented into the combination of the reference patterns, such as shown in Fig.2.12. Here, the weighing values can be seen as the feature of this block.

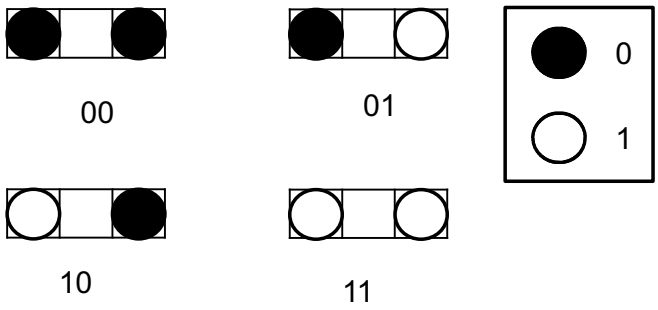


Fig. 2.11: Reference patterns for $P^{'} = 2$ and $h = 1$, the center of each pattern is *pivot*.

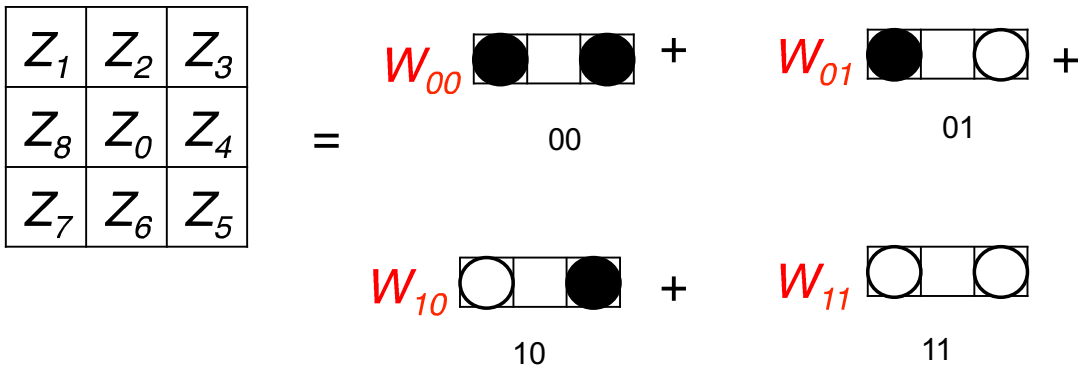


Fig. 2.12: Example of representation of one block.

Fig.2.13 and Fig.2.14 show the features of one block and the related rotated block, respectively. Note, here, just one resolution is used and $\xi = 8$. Fig.2.15 illustrated the features of the same block and the related rotated block by LBP. From these three

figures, we can see clearly that the multi-scans based local patterns are rotation invariant while LBP is not.

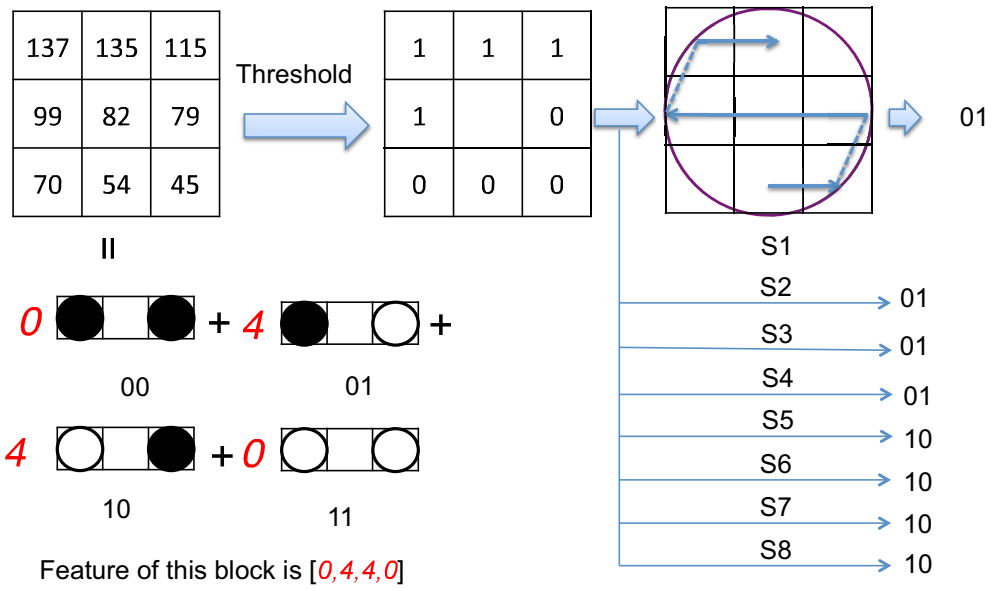


Fig. 2.13: Feature of one block by multi-scans based local patterns.

Fig.2.16 and Fig.2.17 show the comparison of patterns obtained by LBP and our proposed multi-scans based local patterns when the outside factors are changed, such as illumination or noise. From these two examples, we can see that our proposed patterns can keep six same patterns while two are different, but LBP operators are totally changed. Thus, our proposed patterns are robust to the outside factors change.

In a word, after getting the transformed sequence, 1D local patterns(1DLP) is assigned to every pixel by thresholding the P'-neighborhood for every pixel with *pivot* and treating the result as a binary, ternary or quaternion number respectively. Here, the definition of local neighborhood denotes a group of sampling pixels which evenly space on the scan order and center at the point to be calculated. Any interval as well as number of sampling pixels are accepted. If the sampling point can not

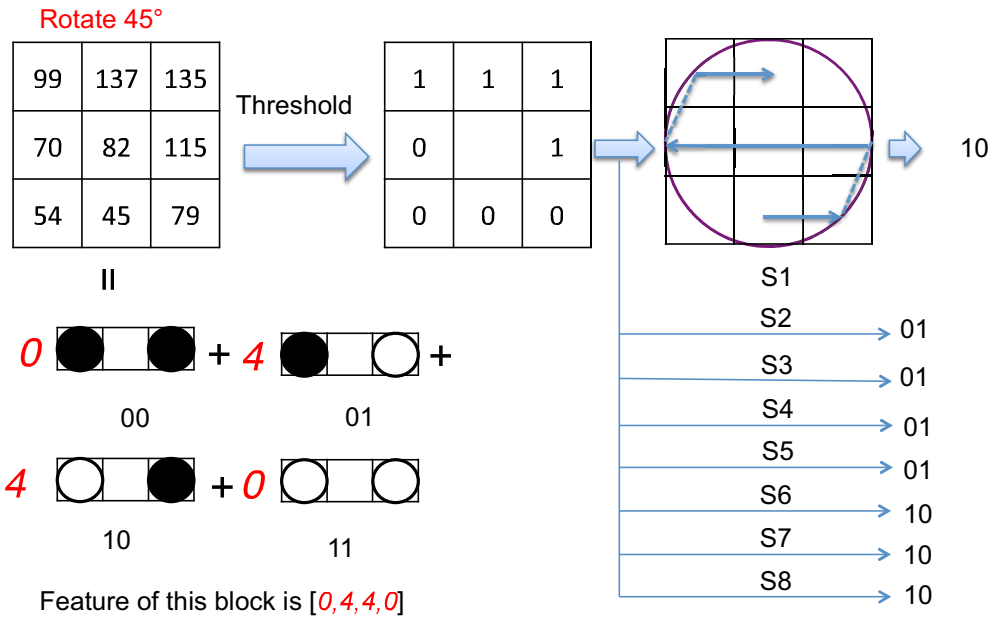


Fig. 2.14: Feature of corresponding rotated block in Fig.2.13 by multi-scans based local patterns.

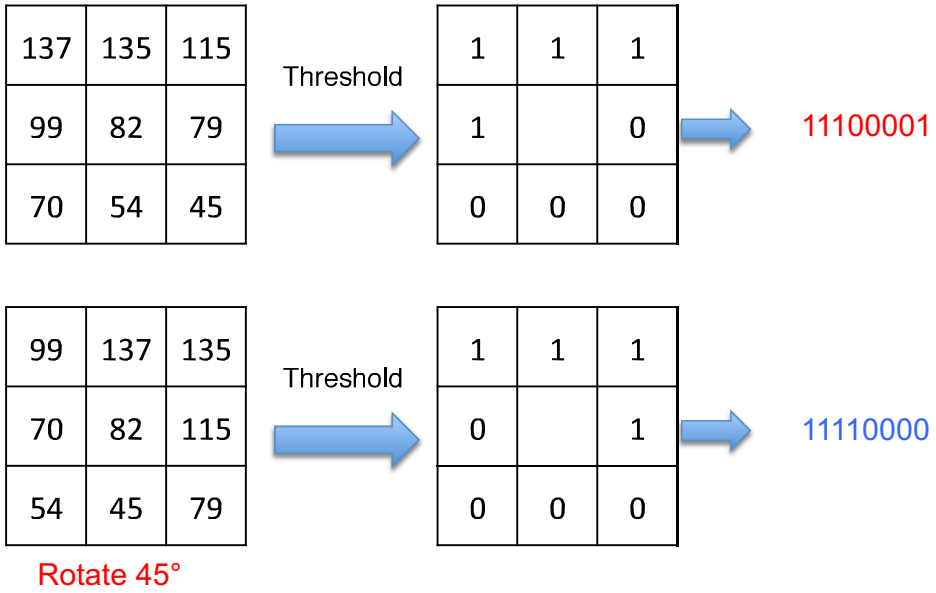


Fig. 2.15: Features of one block and the related rotated block by LBP.

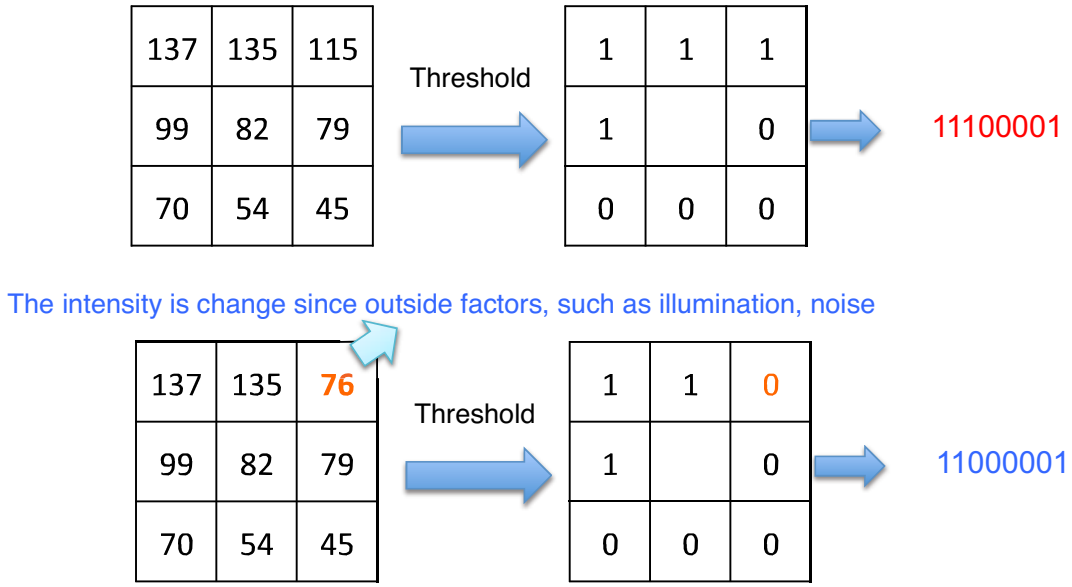


Fig. 2.16: Comparison of LBP when the outside factors change.

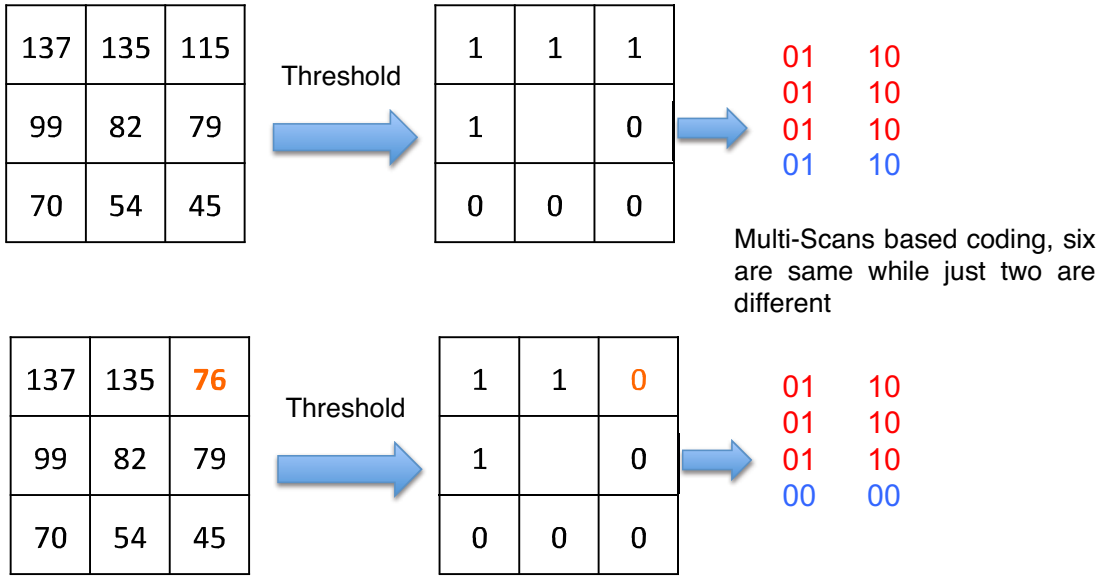


Fig. 2.17: Comparison of multi-scan based local patterns when the outside factors change.

locate at the center of a pixel, simple bilinear interpolation is applied . Fig.2.18 gives an example.

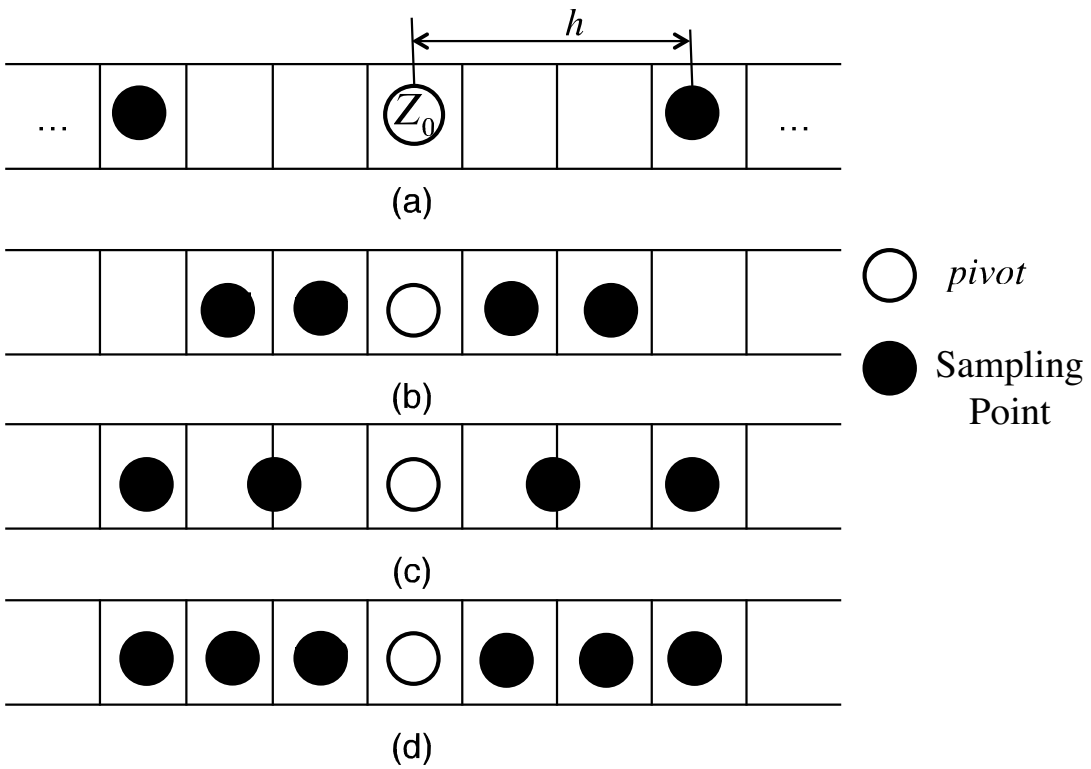


Fig. 2.18: Bilinear interpolation. (a) Subsequence from transformed block. (b) (4,1), (c) (4,1.5), and (d) (6,1) neighborhoods. Black circle without filling is the location of *pivot* and all black circles are the locations of sampling points.

Then, some 1D local patterns called 1D Local Binary Patterns(1DLBP), 1D Local Ternary Patterns(1DLTP), 1D Local Quaternion Patterns(1DLQP) are obtained based on varied encoding rules shown in Eq. (2.1), (2.2), (2.3), respectively.

Images reconstructed by 1DLTP and 1DLQP are smoother than 1DLBP [26]. We can see this point in reconstructed images (Fig.4.3). Fig.2.19 gives an subsequence transformed from a block. The number is the intensity of each element and the one within circle is the *pivot*. Table 2.1 shows an illustration of the three kinds of 1DLP. For each scan, 1DLP can be obtained, and 1D local patterns by multi-scans (1DLPMS) is the combination of them. After getting all patterns for an image, a histogram can be drawn based on the number of patterns, which is marked as feature

descriptor of this image.

...	211	130	150	196	82	...

Fig. 2.19: Subsequence transformed from a block (the number is the intensity of each element).

Table 2.1: 1D local patterns for Fig.2.19 ($P^{'}$=4, h=1, τ=2).

Method	Pattern
1DLBP	1010
1DLTP	2110
1DLQP	3120

Since most shape information in images, such as facial or texture image, is fundamental [6], such as line, edge, multi-scans can capture them effectively and efficiently. For example, S1 and S3 can capture vertical and horizontal line and edge information, S2 and S4 can capture diagonal line, respectively. These fundamental information can be coded by multi-scans accurately and fast while it is less sensitive to outside variations. Also, this point can be seen in Fig.4.3, especially, S3 can get more clear information of eyebrows, eyes and mouth which are the key parts in the facial image.

(a) (b) (c) (d) (e)

Fig. 2.20: Some representative samples for different R, P' and h by 1DLTP and S3 scan. (a) Original image. (b) $R = 1$, $P' = 2$, $h = 1$; (c) $R = 1$, $P' = 4$, $h = 1$; (d) $R = 2$, $P' = 4$, $h = 2$; (e) $R = 2$, $P' = 4$, $h = 3$.

Some representative samples by 1DLTP using S3 scan for different R, P' and h are illustrated in Fig.2.20. From this figure, we can see that all the key parts in faces

are kept clearly. With a little larger R, P' and h, more detailed information of the key parts can be obtained which is very useful for face recognition. More samples can be seen in Fig.4.3, and compared to LBP encoding approach, less noise is kept in our proposed methods.

2.3 Grouped 1D Local Patterns by Multi-Scans (G1DLPMS)

In our proposed 1D local patterns, especially 1DLQP, the parameter $P^{'}$ determines the number of patterns. A large $P^{'}$ produces a long histogram which means lots of useless information will be generated. When the number of sampling points increases, the number of patterns for basic 1D local patterns will become very large, it will be $4^{P^{'}}$ for 1DLQP, as shown in Fig.2.21. Due to this rapid increase, it is difficult to extend them when a large number of sampling points is needed. This limits its applicability.

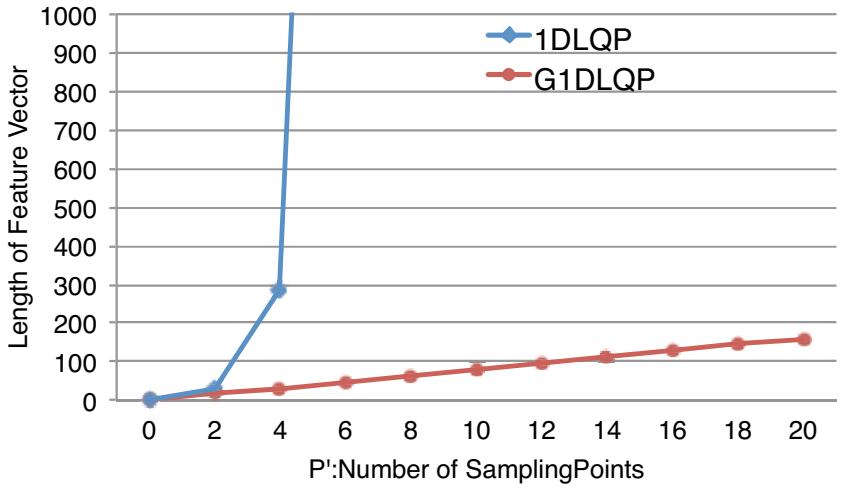


Fig. 2.21: The number of sampling points vs. the number of patterns.

To address this problem, we simplify the descriptor into Grouped 1D Local Patterns (G1DLP) by concatenating 1D local patterns on some groups. Let D_i be the distance between $i-th$ sampling point and *pivot* in the scan sequence and $D_i^{'}$ be

the sorted increasing distance with the rise of i for $i = 1, 2..., P'$. Then, a basic group(called group 1) can be set, where the sampling points with number P'_G are close to the *pivot*. For symmetry, the distribution on both side of *pivot* is same. Thus one basic distance set $D(1) = \{D'_1, D'_2, ..., D'_{P'_G}\}$ which contains top $P'_G - th$ smallest distance can be obtained. Some other distance set $D(t)$ is defined as Eq. (2.6)

$$\begin{aligned} D(t) = \{&D'_1 + (t-1) \times D_M, ..., D'_{P'_G} + (t-1) \times D_M\}, \\ &where \quad D_M = \max\{D'_1, D'_2, ..., D'_{P'_G}\}, \\ &and \quad t = 1, 2, ... \end{aligned} \tag{2.6}$$

The $i-th$ sampling point can be set to group t if and only if $D_i \in D(t)$. Fig.2.22 shows a simple example.

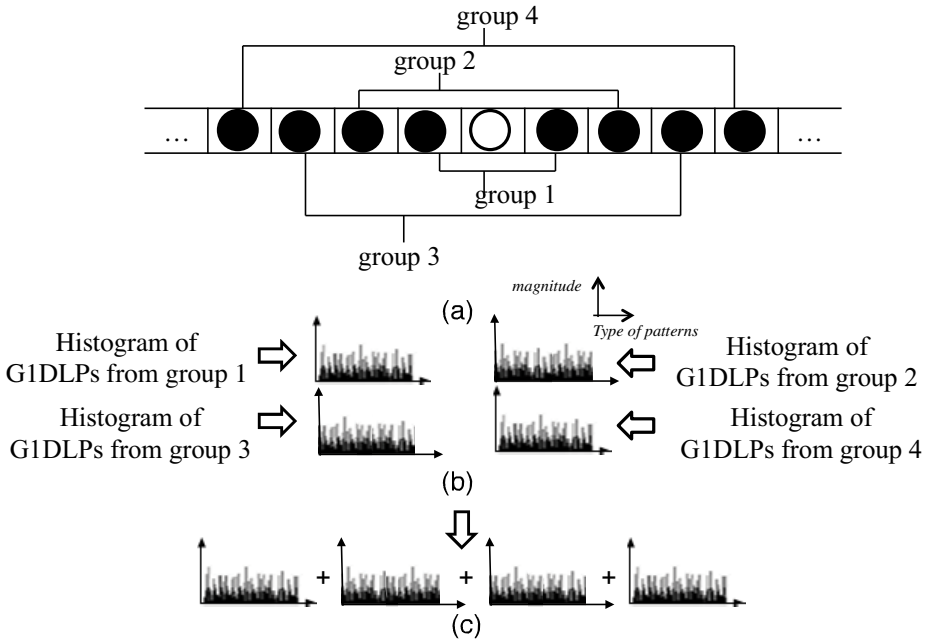


Fig. 2.22: Grouped 1D local patterns. (a) Subsequence transformed from a block. (b)1D local patterns histogram from each group. (c) Concatenated histogram.

The 1D local patterns are extracted from each group to concatenate grouped

1D local patterns (G1DLP), such as grouped 1DLBP (G1DLBP), grouped 1DLTP (G1DLTP), grouped 1DLQP(G1DLQP). Same as 1DLPMS, for each scan, G1DLP can be obtained, and grouped 1D local patterns by multi-scans (G1DLPMS) is the combination of them. For the whole pixels, the statistics of these groups can be got and concatenated to a single histogram (Fig.2.22). So the length of feature vector is linear to the number of sampling points(shown in Fig.2.21 with $P'_G = 2$).

2.4 Evaluation

In this section, our proposed methods will be simply evaluated on the general database which is obtained from Internet (More detailed evaluations will be done in chapter 4 and chapter 5). On this database, 50 categories are selected including buildings, scenes, faces, cars, textures and so on. Each category contains 10 samples, one sample is the original one and the others are alternative ones with the variations of scale, rotation, illumination or noise. Some samples of this database are shown in Fig.2.23. And we will evaluate our proposed approaches with different variation of scale, rotation, illumination and combination of them. In the evaluation of each special variation, just one original image is treated as gallery and the corresponding variation ones are used for probe. -B, -T, -Q means different encoding rule is used under Eqs. (2.1), (2.2), (2.3), respectively.

2.4.1 Scale

The first experiment shows the performance of scale variation. Here, several downsampling factors D and upsampling factors U sets are used. And $D^1 \in \{2\}$, $D^2 \in \{2,3\}$, $D^3 \in \{2,3,4\}$, $U^1 \in \{2\}$, $U^2 \in \{2,3\}$, $U^3 \in \{2,3,4\}$. Table 2.2 illustrated the accuracy with different (D,U) sets and R ($\Delta\theta = 45^o$). From this table, we can see that with larger downsampling and upsampling factors, a little higher accuracy can be obtained. Since a block with larger R can get more information, the performance

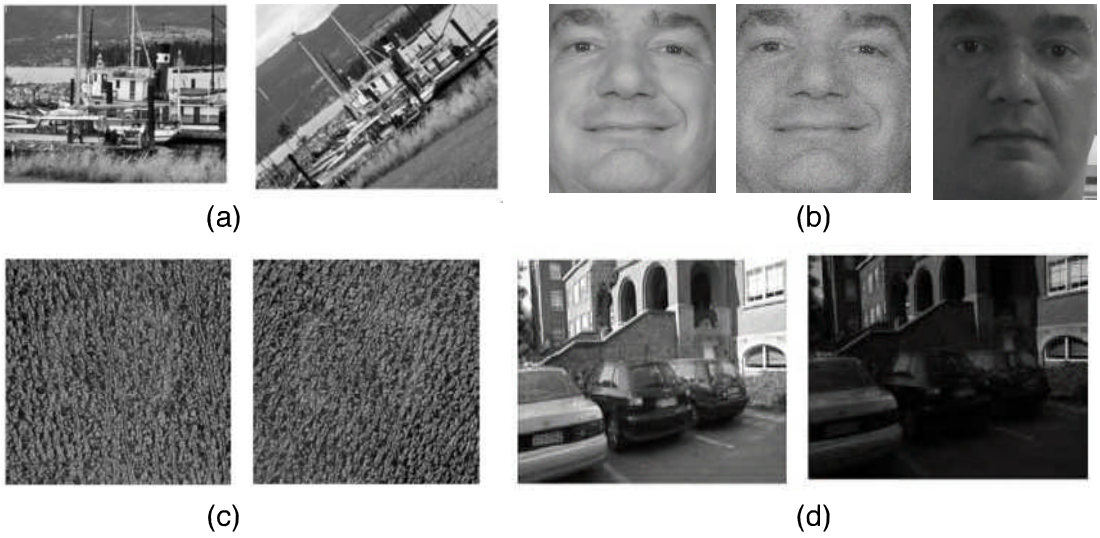


Fig. 2.23: Samples of the evaluated database. (a) Variation of scale and rotation, (b) variation of noise and illumination, (c) variation of rotation, (d) variation of illumination.

is also a little increased. But, on the other side, larger D, U and R also take time for processing. Thus, in order to take the balance between the accuracy and time, sets D^2, U^2 and radius $R = 2$ are used by default.

Table 2.3 shows final recognition rate among several methods with the variation of scale on this database. From the table, we illustrate that our proposed methods are comparable with famous scale invariant features, such as SIFT and SURF.

Table 2.2: Accuracy (%) on the scale variation sub-database with different resolutions and R, 1DLPMS-T approach is used.

(D,U) \ R	1	2	3
(D^1, U^1)	90.3	90.3	90.5
(D^1, U^2)	90.9	91.5	91.7
(D^1, U^3)	90.9	91.7	91.7
(D^2, U^1)	91.1	91.7	91.7
(D^2, U^2)	91.5	92.1	92.1
(D^2, U^3)	91.7	92.1	92.3
(D^3, U^1)	91.1	92.1	92.1
(D^3, U^2)	91.7	92.3	92.1
(D^3, U^3)	91.9	92.1	92.3

Table 2.3: Accuracy on the sub-database with variation of scale.

Method	Accuracy (%)
LBP [6]	89.7
SIFT [3]	92.5
SURF [4]	90.7
ALBP [7]	89.5
LDP [9]	92.3
LGBP [10]	93.1
1DLPMS-B	90.3
1DLPMS-T	92.1
1DLPMS-Q	92.6

2.4.2 Rotation

The second experiment shows the influence of rotation on the related methods. Table 2.4 illustrates the accuracy on the rotation variation sub-database with different $\Delta\theta$ and R. From the table, we see that a little small $\Delta\theta$ and larger R can get higher result, since more orientations of the sequence and more information of the image can be encoded. But the time complexity is also increased. Thus, by taking the trade-off the accuracy and time, $\Delta\theta = 45^0$ is set by default in the following experiments.

Table 2.4: Accuracy (%) on the rotation variation sub-database with different $\Delta\theta$ and R, 1DLPMS-T approach is used.

$\Delta\theta$ \ R	1	2	3
15^o	90.1	90.7	90.7
30^o	90.3	90.5	90.7
45^o	90.1	90.5	90.5
60^o	88.5	89.9	90.3
90^o	83.1	84.9	84.5

Table 2.5 shows final recognition rate among some methods under rotation variation. From the table, we illustrate that our proposed methods are also comparable with famous rotation invariant features, such as SIFT, SURF and LGBP. While LBP is very sensitive to the rotation changes.

Table 2.5: Accuracy on the sub-database with variation of rotation.

Method	Accuracy (%)
LBP [6]	13.7
SIFT [3]	87.9
SURF [4]	85.3
ALBP [7]	87.4
LDP [9]	88.5
LGBP [10]	89.8
1DLPMS-B	89.3
1DLPMS-T	90.5
1DLPMS-Q	91.9

2.4.3 Illumination

This third experiment shows the illumination effects of the methods. Table 2.6 shows final recognition rate. From this table, we can recognize that our proposed methods are much better than some famous algorithms. Since LBP, SIFT and SURF are robust to monotonic lighting variety but very sensitive to non-monotonic illumination changes. Our proposed adaptive thresholding strategy and multi-scans encoding can tackle this kind of problem much effectively.

Table 2.6: Accuracy on the sub-database with variation of illumination.

Method	Accuracy (%)
LBP [6]	47.7
SIFT [3]	37.5
SURF [4]	45.2
ALBP [7]	57.1
LDP [9]	83.8
LGBP [10]	89.6
1DLPMS-B	81.3
1DLPMS-T	91.2
1DLPMS-Q	93.5

2.4.4 Total Accuracy

In this evaluation, the recognition performance of our proposed algorithms and some famous feature extraction methods is compared on the total database. Two samples

of each category are randomly chosen as gallery (training images), and the left ones are applied for probe (testing images). In our study, we apply 3 times to randomly select the training sets and in final the average recognition rate is recorded.

The final results are concluded in Table 2.7. From this table, we can illustrate that our proposed local patterns can deal with variations of scale, rotation and illumination more robust than traditional approaches.

Table 2.7: Total accuracy on the whole database.

Method	Accuracy (%)
LBP [6]	56.8
SIFT [3]	69.5
SURF [4]	72.7
ALBP [7]	83.1
LDP [9]	88.5
LGBP [10]	89.3
1DLPMS-B	85.3
1DLPMS-T	91.9
1DLPMS-Q	94.2

2.5 Conclusions

In this chapter, based on the proposed multi-scans encoding model, a novel feature descriptor for presenting the local patterns is developed and its simplifications and extensions are also considered. Our proposed method - 1DLPMS can capture more spatial information and key parts clearly with less time complexity as well as it is scale, rotation and illumination invariant. In order to get computational simpler and easier to extended approach, just the co-occurrences of local patterns in some groups are taken into account. Experiments on a general database show that our proposed methods have better performance than traditional ones. Especially, based on the separated evaluation on illumination, we can clearly recognize the effectiveness of our proposed multi-scans encoding model based feature descriptors.

3

Invariant Local Curvelet Feature Descriptors by Multi-Scans

3.1 Introduction

Our proposed methods in this chapter also belong to local feature group. General speaking, there are many directions to be researched in this area. In chapter 2, firstly, multi-scans are used to encode the neighborhood of the patterns and concluded that it can reserve more spatial information with lower cost than the traditional neighborhood encoding, such as circle neighborhood based encoding. Secondly, in order to improve the robustness of local patterns, not just binary, ternary and quaternion number are used to encode the neighborhood structure, which is more robust to outside factor change. However, the multi-scans neighborhood based encoding is also based on gray-level intensity which is sensitive to the outside factors change, and the type of patterns are predefined which are not patten-specific. Thus, in this chapter, firstly, the gray-scale intensity is replaced by Curvelet feature to learn some local patterns. The advantage of this kind of transformation is that it is expected to capture salient and effective properties, for instance spatial localization, spatial

frequency and orientation selectivity, thus these local patterns learned by Curvelet feature become more superior to the ones encoded by gray-level features. Secondly, clustering method is used to learn a pattern-specific codebook from the training set, which is more suitable for pattern perception tasks. One point should be noted is that in this chapter, for simplicity, just binary number is used for encoding the local patterns. But it is easy to extend to use ternary and quaternion number for encoding.

To select more compact and discriminative features and solve the outliers or occlusion problems as well as further reduce the feature dimensions, one novel Linear Discriminate Analysis (LDA) approaches is proposed, which is based on L1-norm optimization. LDA based on a novel L1-norm optimization method has several advantages: firstly, it can deal with the outliers or occlusion problem effectively. Secondly, it is invariant to rotations. Thirdly, it has no Small Sample Size problem, which usually exists in traditional L2-norm based LDA methods.

Evaluation based several databases and a reconstruction problem will be used to estimate the effectiveness of our proposed approaches.

The remainder of this chapter is organized as follows: Curvelet Transform will be introduced in section 2. Image representation with local Curvelet feature descriptors will be described in section 3 and section 4. LDA-L1 will be proposed in section 5 and evaluation will be estimated in section 6. Finally, conclusions are discussed in section 7.

3.2 Curvelet Transform

Curvelet is a wonderful non-adaptive approach to represent multi-scale object, which can deal with useful curved edges in image processing and pattern recognition. As shown in Ref. [29], compared to traditional wavelet, Curvelet can represent smother edge and curve information by fewer coefficients. Both Curvelet and wavelet are the methods of the multi-scale geometric transformation. In general, they can produce

multi-scale and multi-orientation features around the image by pyramid construction. But Curvelet [29] is superior over the wavelet as follows: 1. It can represent the edges or curves of the objects sparsely and optimally. 2. For ill-posed problems, it can reconstruct the image optimally. 3. It can also represent the wave propagators sparsely and optimally.

Recently, Fast Discrete Curvelet Transform (FDCT) [29] is a novel and improved version of original Curvelet Transform (CT). Compared to the traditional CT that is based on ridgelets, the FDCT is faster, simpler with less redundancy. In Ref. [29] , Candes et al. proposed two versions of FDCT: one is applied with Unequally-Spaced Fast Fourier Transforms (USFFT), while the other is based on a wrapping function. The most difference between these two versions is that how to select the spatial grid to translate Curvelet in each angle and scale. No matter which version, in final, a spatial location parameter, an orientation parameter and digital Curvelet coefficients are returned. Compared to USFFT, wrapping function-based approach, which uses special selected Fourier samples, is easier to understand and implement. Thus, in our study, wrapping-based CT is applied.

In detail, wrapping-based CT uses a 2D image for input in the Cartesian space with the array $f[m, n]$, where $0 \leq m < M$, $0 \leq n < M'$ where M and M' are the dimensions of the array. As shown in Eq. (3.1), the outputs are a set of Curvelet coefficients $c^D(u, v, k_1, k_2)$ indexed by an orientation v ,a scale u, and spatial location parameters k_1 and k_2.

$$c^D(u, v, k_1, k_2) = \sum_{m,n} f[m, n]\varphi^D_{u,v,k_1,k_2}[m, n]. \tag{3.1}$$

Each φ^D_{u,v,k_1,k_2} is a digital Curvelet waveform, superscript D stands for "digital" [29]. Generally, CT can extract more effective and efficient curved edges in frequency domain. And wrapping-based CT is a multi-scale pyramid that consists of some sub-

bands at different scales, orientations and locations. Curvelet looks like so fine and a needle shaped element in high frequency level, while it the low frequency level it seems non-directional coarse elements.

In our study, the image is decomposed into coarse, detail and fine coefficients and some reconstructed images (here, facial image is used) including coarse layer, two detail layers and one fine layer are illustrated in Fig.3.1. F

.3.2. Here, CurveLab 2.1.2 which is available at [30] is used.

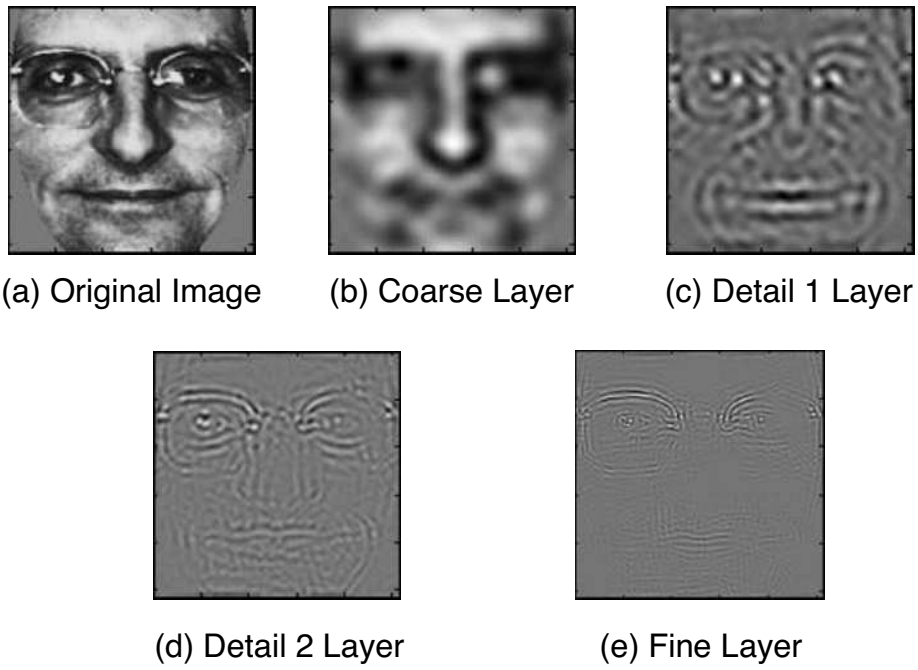


Fig. 3.1: Reconstructed images from different Curvelet coefficients.

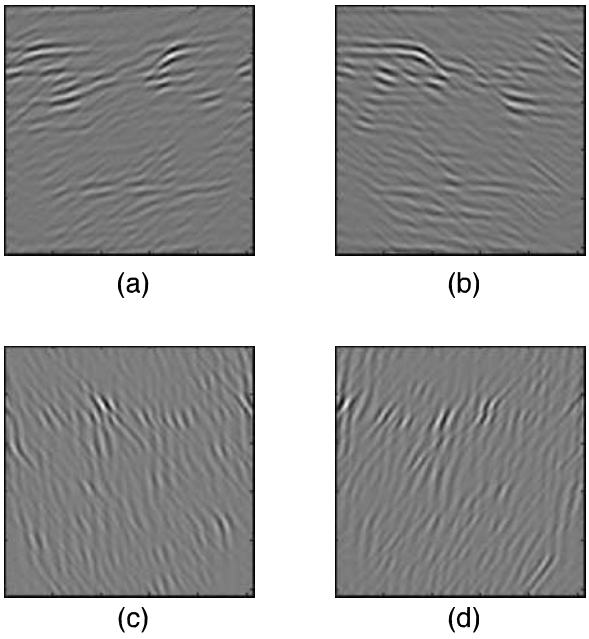


Fig. 3.2: Reconstructed images from detail 2 layer by four orientations.

3.3 Image Representation with Local Curvelet Binary Pattern by Multi-Scans (LCBPMS)

Based on the Curvelet transform, the overall framework of our proposed representation approached called Local Curvelet Binary Pattern by Multi-Scans (LCBPMS) is illustrated in Fig.3.3. In this method, an image can be reconstructed as a histogram sequence in the following steps: First, the original image is transformed to obtain multiple Curvelet Pictures (CPs) in frequency domain by applying Curvelet filters; Second, each CP is converted to LCBPMS map by 1DLPMS operator; Third, each LCBPMS map is further separated into several regions, and the histogram of each region is calculated; Fourth, all the LCBPMS histograms of the LCBPMS maps are then concatenated into construct the final histogram sequence which is defined as the feature descriptor of the image.

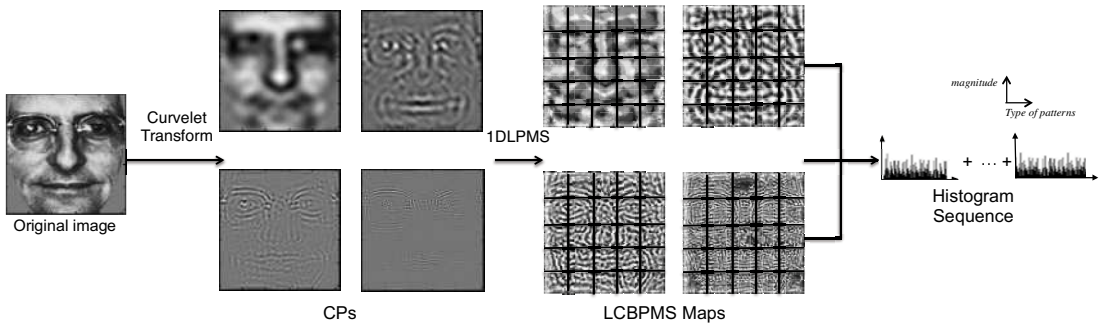


Fig. 3.3: The framework of proposed LCBPMS for image representing phase.

3.3.1 LCBPMS

LCBPMS can be regarded as a combination of Curvelet feature and 1DLPMS-B operator. In this study, just binary thresholding strategy is considered for simplification. Of course, it is easy to extend to use ternary and quaternion number for encoding.

3.3.2 LCBPMS Histogram Sequence

Since histogram statistical presentation is successfully used in local feature areas, LCBPMS Histogram Sequence is considered which can be generalized from LCBPMS maps in our study. First, each LCBPMS map is spatially divided into B regions. And then the histogram of every region is extracted. In final, all of the histograms extracted from the regions in the LCBPMS map can be concatenated together into a final histogram sequence for LCBPMS map $L_a(.)$ ($a = 0, 1, 2, 3$ corresponding to coarse layer, detail 1 layer, detail 2 layer and fine layer in our implementation, respectively), that can be formally described as following:

$$H_a(.) = [H_{a,0}, H_{a,1}, ..., H_{a,B-1}]. \tag{3.2}$$

Here $H_{a,i}(i = 0, 1, .., B - 1)$ stands the histogram of the i-th region of LCBPMS map $L_a(.)$ and is formulated as follows:

$$H_{a,i}(j) = \sum_{Z} B\{L_a(Z) = j\}, j = 0, 1, ..., K-1, \tag{3.3}$$

where j is the index of the patterns, K denotes the number of the patterns, Z is the pixel location in LCBPMS map and $B\{.\}$ is defined as follows:

$$B\{A\} = \begin{cases} 1, & \text{if A is true,} \\ 0, & \text{others.} \end{cases} \tag{3.4}$$

In final, one image Img can be encoded as a histogram sequence, which concatenates all the region based histograms at all LCBPMS pattern maps with different Q layer:

$$H(Img) = [H_0(Img), H_1(Img), H_2(Img), ..., H_Q(Img)]. \tag{3.5}$$

3.4 Image Representation by Learned Local Curvelet Pattern (LLCP)

Different from the patterns used in LCBPMS which are predefined in the same way for special object and non-special object images, several particular codebook patterns are learned in LLCP (shown in Fig.3.4). First, the special object images are sampled into many patches and then all the patches are cluttered into K patterns by random-projection tree [31]. Generally speaking, image encoding by LLCP can be grouped into learning stage and representing stage. In the learning stage, several particular codebooks are constructed while images are encoded in the representing stage.

In the learning phase (shown in Fig.3.5), every sample in the training group is reconstructed to CPs with different Curvelet coefficients. Then based on all CPs in the same la l the patches. At last, by using random-projection tree clustering method in every patch set, LLCP learned codebooks can be constituted. Thus, Q LLCP codebooks can be obtained with Q

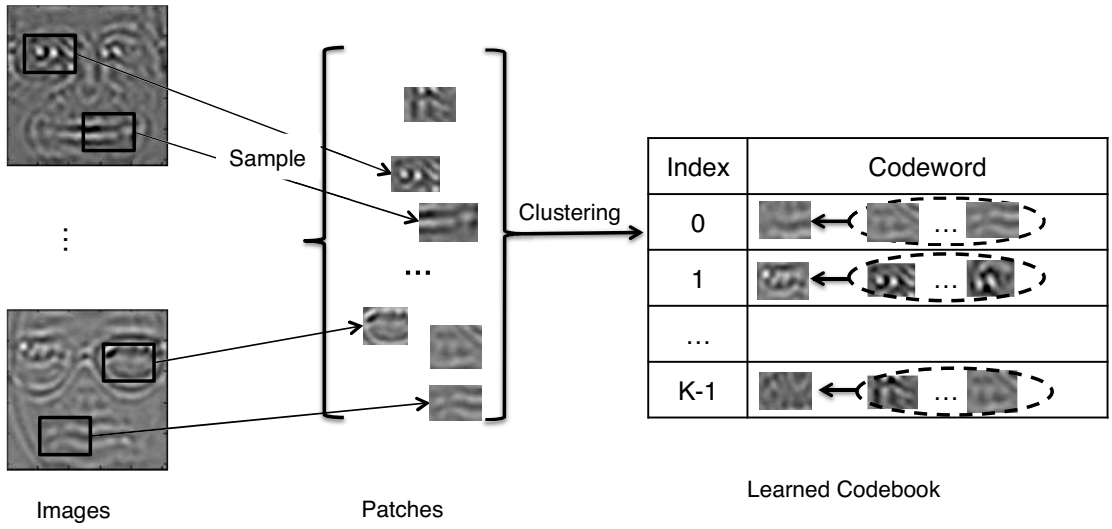


Fig. 3.4: The process of learning a particular codebook.

layers.

One point should be explained in our implementation procedure is that the sampling methods used are illustrated as Fig.3.6, which called sample 1 and sample 2, respectively.

In the representing phase (shown in Fig.3.7), first, each image is reconstructed to several CPs with different Curvelet coefficients. Then each CP can be encoded into T pattern maps through projecting its patches into the related LLCP codebooks pixel by pixel (Note: here the intensity of mapped image is corresponding to the type of learned patterns, and also the same intensity in different LLCP maps stands for different type of patterns since the codebook is different. More important one is that one CP can be encoded by T LLCP maps and the reason will be explained later). Thus, totally $Q \times T$ LLCP maps can be obtained. In final, the pattern maps are separated into some regions spatially and all the histograms of the regions are concatenated together to construct an enhanced histogram sequence that is considered to represent the input image.

Here, Euclidean distance $D_e(L_{a,t}(Z), j) = ||P_{L_{a,t}(Z)} - C_j||_2$ is used to calculate

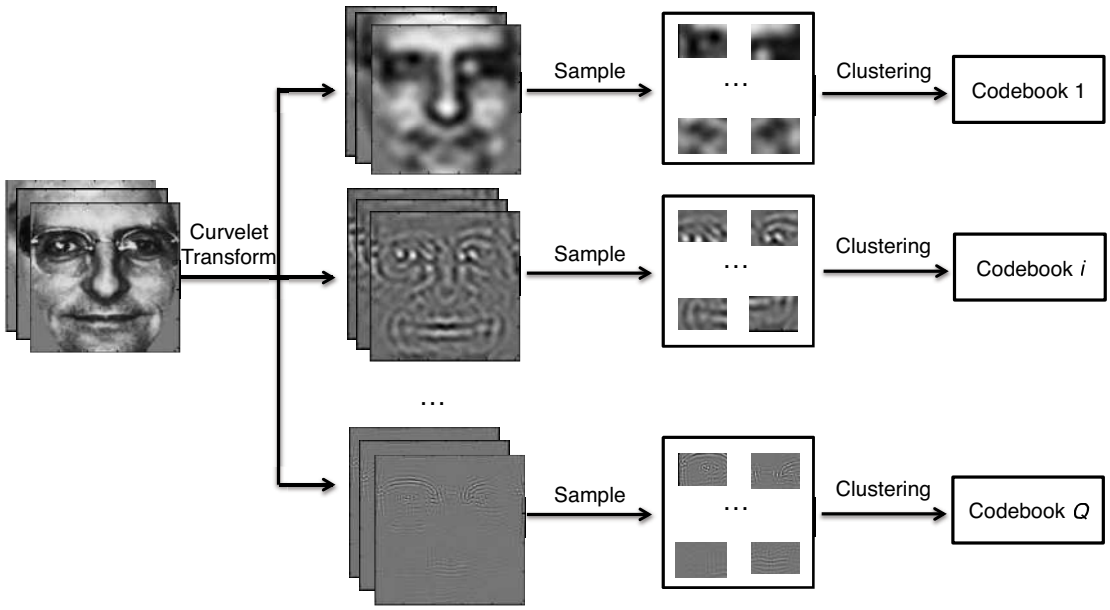


Fig. 3.5: The learning phase for image encoding by LLCP.

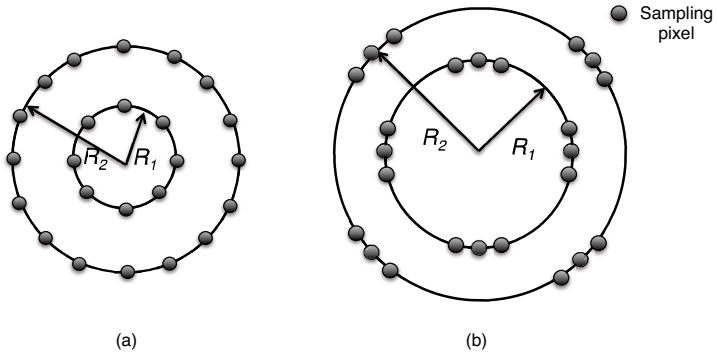


Fig. 3.6: Sampling methods used in our implementation. (a) Sample 1: $R_1 = 1$, $R_2 = 2$; (b) Sample 2: $R_1 = 4$, $R_2 = 7$.

the similarity score between the input patch and the codewords ($L_{a,t}(.)$ stands for the map in layer a where the encoded pattern is t-th smallest distance codeword ($t = 1, 2, .., T$), C_j is the codeword for pattern j and $P_{L_{a,t}(Z)}$ is the input patch at position Z in $L_{a,t}(.)$). One point should be noted during representing phase is that the input patch can match top T-th smallest distance codewords in the codebook. The reason is that the input patch is very closed to the top T-th smallest distance

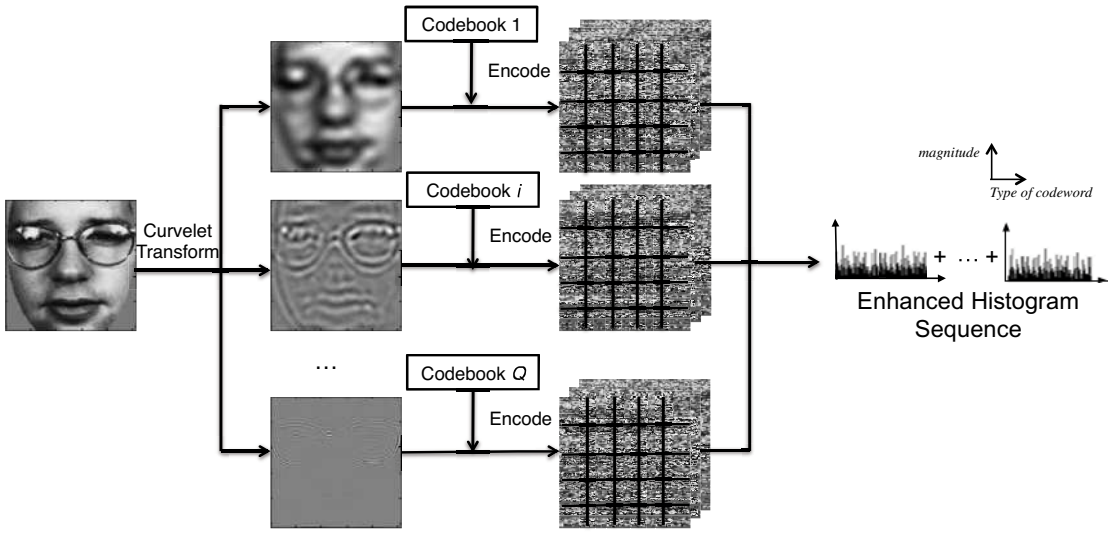


Fig. 3.7: The representing phase for image encoding by LLCP.

codewords. And at most time, we can not judge which one is the best matching codeword. One example is shown in Fig.3.8 where codeword 0 and codeword 2 have the same similarity score, so in such case we can not just select one and ignore the other. That is, the traditional one to one mapping is not useful in this situation and multi-mapping strategy is applied to overcome the problem.

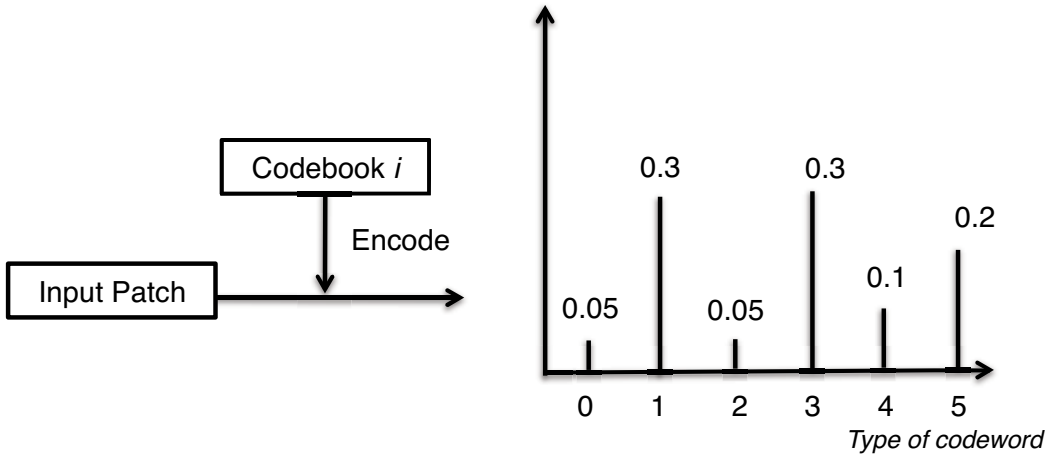


Fig. 3.8: Example of getting similarity scores between input patch and codewords, here $K = 6$.

So when histogram encoded by LLCP is computed, Eq. 3.3 should be changed as following

$$H_{a,i}(j) = \sum_t \sum_Z B\{D_e(L_{a,t}(Z), j) \leq \beta \min_k D_e(L_{a,1}(Z), k), \\ k = 0, 1, ..., K-1\}, \tag{3.6}$$

where $\beta(\beta \geq 1)$ is a predefined real threshold value and k is the index of codeword.

3.5 L1-norm based Linear Discriminant Analysis

In this section, one novel and effective feature projection method will be proposed based on L1-norm.

3.5.1 Related Work

In many data analysis issues, data measured or observed should be located in a low but powerful subspace instead of the original high dimension. Since this kind of subspace has lots of significant applications in image processing and pattern recognition, such as motion estimation [32], face recognition [33]. Especially, in face recognition system, the feature descriptor of faces is always located in high dimension and it would take times to do classification, so dimension reduction is a key and powerful step in face recognition task. Among these subspace methods; linear discriminant analysis (LDA) [34] is one of the widely used approaches. LDA aims to obtain a set of projections that maximize the ratio of the between-class (S_w) distance to the within-class (S_w) distance. These projections consist of a low but effective dimensional linear subspace where the data structure, such as face features, in the original input dimension can be efficient and powerful represented.

The classical LDA [34] [16] was proposed to find an optimal discriminant but low-dimensional subspace to maximize the S_b separability of the data samples and their S_w compactness . However, in many cases, the classical LDA has the issue that called

Small Sample Size (3S) problem, since it needs any of scatter matrices (S_w) should be nonsingular to calculate S_w^{-1}. But unfortunately, the size of the training set seems much smaller than the dimension of the feature space in lots of applications, such as face recognition. In recent decades, some LDA extensions have been presented to overcome such 3S problem. The most famous one was called Fisherface [35] , which first applies PCA to decrease the original data and then uses LDA to extract the discriminative information. However, this approach may lose key discriminative structure information in the PCA step for further classification process. In [36], null-space linear discriminant analysis (NLDA) was proposed, which projects all the samples into a null space of S_w and then extracts discriminative information. In [37], direct linear discriminant analysis (DLDA) was studied to extract the discriminative information from the null space of S_w matrix, achieved by diagonalizing first S_b then diagonalizing S_w. A common disadvantage of the previous approaches is that the discriminative vectors are solved by a single data sample substructure rather than the whole data samples space. Thus, some powerful discriminative structure [38] [39] will be ignored to a certain extent. Thus, in [38], the researchers proposed a dual-space LDA (DSLDA) approach that uses total advantage of the discriminative structure in gallery. The key merit of the DSLDA approach is that the full data samples are separated to two complementary substructures, such as a range dimension for the within-class scatter matrix as well as the complementary one, and then the discriminative features in every subspace are solved. However, the time cost of DSLDA is very high so that it is not suitable to real-time applications. Ref. [39] proposed a complete kernel fisher discriminant analysis (CKFD) method to deal with discriminant analysis in both scatters.

However, Frobenius norm (L2-norm), which is sensitivity to the presence of outliers or noise, is used in all the above LDA approaches to measure S_b and S_w distances. Thus, the process of training maybe dominated by the outliers or occlusion

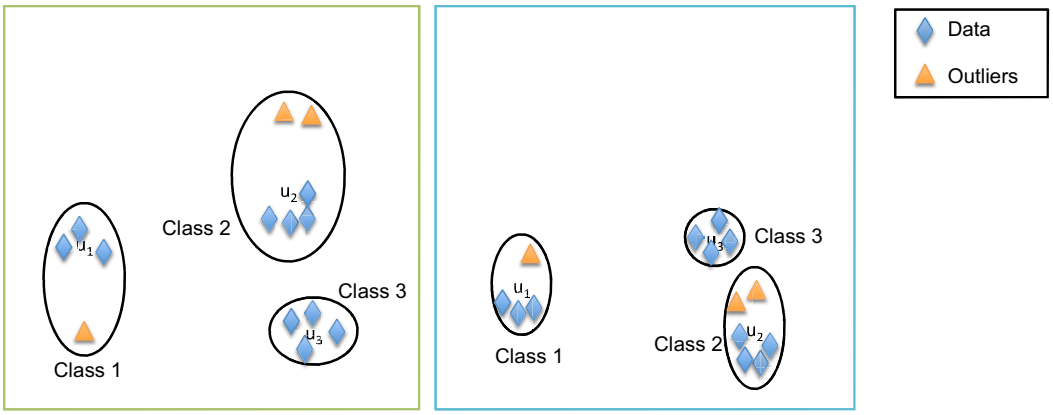
since the S_b or S_w measurement is computed by the sum of squared distances. Inspired by [21], [22] to decrease the influence of outliers and occlusion problem, a novel L1-norm based linear discriminant analysis called LDA-L1 is proposed for robust discriminant analysis (and also we noted the traditional LDA based on L2-norm as LDA-L2 in the following). Let $Z = (z_1, z_2, ..., z_n)$ be n data points in d-dimensional space. In matrix form $Z = (z_{ji})$, index j sum over spatial dimensions, $j = 1, 2, ..., d$ and index i sum over data points, $i = 1, 2, ..., n$. L2-norm is defined as

$$||Z||_{L2} = \left(\sum_{i=1}^{n} \sum_{j=1}^{d} z_{ji}^2 \right)^{\frac{1}{2}}, \tag{3.7}$$

and L1-norm is defined as

$$||Z||_{L1} = \sum_{i=1}^{n} \sum_{j=1}^{d} |z_{ji}|. \tag{3.8}$$

Fig.3.9 illustrates the distribution of the features in the original and projected space



(a) Original feature space (b) Projected feature space by LDA-L2

Fig. 3.9: Blue diamonds stand for data and yellow triangles are outliers. (a) The feature distribution of three classes with outliers in original feature space. (b) The feature distribution of three classes and outliers in the projected feature space by LDA-L2.

by LDA-L2. We can see that the data and outliers cannot be separated in both

feature spaces.

Compared to L2-norm, L1-norm is more robust to outliers since no square operation is applied to calculate the distance. For example, in Fig.3.10, ten data samples as well as two outliers are illustrated to fit a 1D subspace line. The results are shown by the solid line and dash line for L1-norm and L2-norm, respectively. We can see that L1-norm obtains correct result but L2-norm obtains erroneous line fitting.

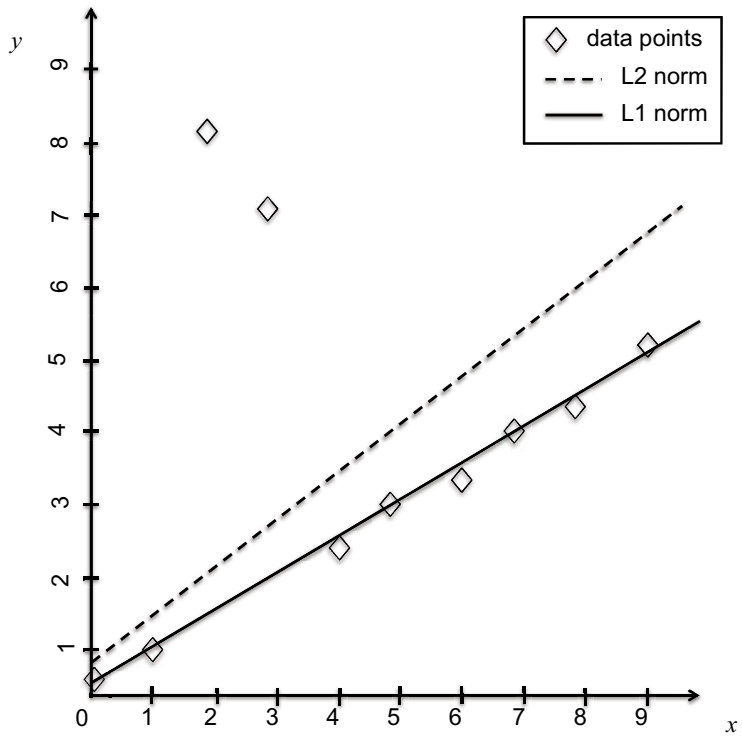


Fig. 3.10: Fit a line for 10 given data points. The two data located on upper-left are outliers.

Recently, in order to solve the outliers problem, Li [23] proposed rotation invariant L1-norm (notated as R1-norm) based linear discriminant analysis (we call it LDA-R1 in the following). The R1-norm is defined as

$$||Z||_{R1} = \sum_{i=1}^{n} \left(\sum_{j=1}^{d} z_{ji}^2 \right)^{\frac{1}{2}}. \tag{3.9}$$

Here, R1-norm is calculated by the summation of samples without being squared. Therefore, the R1-norm should be more robust to outliers compared to L2-norm. However, LDA-R1 takes lots of cost to achieve the final convergence for a large dimensional input space. In this chapter, instead of maximizing the variance computed by L2-norm, L1-norm based linear discriminant analysis is proposed. Based on the reports in [21], our proposed method is expected less sensitive to outliers than L2-norm and R1-norm based approaches, and it also does not have 3S issue since it does not want to compute the inverse of scatter matrix S_w^{-1}. In addition, our proposed method is simple and easy to implement.

3.5.2 Problem Formulation

Assume we have a set of samples $X = \{\{x_i^l\}_{i=1}^{N_l}\}_{l=1}^{C} \in \mathbb{R}^{d\times n}$. N_l of which belong to class ω_l $(l = 1, 2, ..., C)$, where n is the number of samples while d is the dimension of the original input space, and $n = \sum_{l=1}^{C} N_l$.

In LDA-L2, the objective is to seek t projections $Y = \{\{y_i^l\}_{i=1}^{N_l}\}_{l=1}^{C} \in \mathbb{R}^{t\times n}$ by means of t linear transformation vectors $w_i \in \mathbb{R}^{d\times 1}$, which can be arranged by columns into a projection matrix $W = [w_1, w_2, ..., w_t] \in \mathbb{R}^{d\times t}$, which embeds the original d dimension into t dimension vector space such that $t < d$. Let S_b be the between-class scatter matrix, and S_w be the within-class scatter matrix. Therefore, the between-class and within-class distances can be, respectively, formulated as:

$$S_b = \sum_{l=1}^{C} (m_l - m)(m_l - m)^T, \tag{3.10}$$

$$S_w = \sum_{l=1}^{C} \sum_{i=1}^{N_l} (x_i^l - m_l)(x_i^l - m_l)^T, \tag{3.11}$$

where $m_l = (1/N_l) \sum_{i=1}^{N_l} x_i^l$ is the mean of the samples that are assigned into the l-th

class, and $m = (1/n)\sum_{l=1}^{C} N_l m_l$ is global mean of all the samples. The purpose of LDA-L2 is to obtain an optimal transformation W by maximizing the ratio of $\mathrm{Tr}(S_b)$ and $\mathrm{Tr}(S_w)$ as the following problem

$$\max_{W} J_{L2} = \max_{W} \frac{\mathrm{Tr}(S_b)}{\mathrm{Tr}(S_w)} = \frac{W^T S_b W}{W^T S_w W}. \tag{3.12}$$

It is well known that the L2-norm is not robust to outliers and R1-norm approach was presented to solve this problem [23]. Thus, the issue can be illustrated to find W which maximizes the following objective function:

$$\begin{aligned} \max_{W} J_{R_1} = (1-\alpha)\sum_{l=1}^{C} N_l \sqrt{||W^T(m_l - m)||^2} - \\ \alpha \sum_{l=1}^{C} \sum_{i=1}^{N_l} \sqrt{||W^T(x_i^l - m_l)||^2}. \end{aligned} \tag{3.13}$$

Here, the parameter α is a trade-off predefined coefficient such that $0 < \alpha < 1$. However, it takes lots of cost to achieve the final convergence for a large input dimension space, and also it has null space problem, which appears very often in face recognition application. In this chapter, we want to maximize the L1 criteria by the L1 norm as follows at the feature dimension

$$\begin{aligned} \max_{W} J_{L_1} &= \max_{W} \frac{\sum_{l=1}^{C} N_l ||W^T(m_l - m)||_{L1}}{\sum_{l=1}^{C}\sum_{i=1}^{N_l} ||W^T(x_i^l - m_l)||_{L1}} \\ &= \max_{W} \frac{\sum_{l=1}^{C} N_l |W^T(m_l - m)|}{\sum_{l=1}^{C}\sum_{i=1}^{N_l} |W^T(x_i^l - m_l)|}. \end{aligned} \tag{3.14}$$

The solution of Eq.(3.14) is expected to invariant to rotation since the maximization is based on the feature space that should be less sensitive to the outliers compared to both L2-norm and R1-norm solution. Moreover, no Small Sample Size problem will be occurred as described above.

3.5.3 L1-norm based Linear Discriminant Analysis

Generally, classical LDA used L2-norm (Frobenius norm) to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. However, L2-norm is sensitive to outliers and is not satisfied for robust discriminant analysis. For the sake of addressing this key issue, L1-norm is applied into the objective function and we call this novel approach as L1-norm based Linear Discriminant Analysis (LDA-L1).

Suppose that $W = [w_1, w_2, ..., w_t] \in \mathbb{R}^{d\times t}$ is the orthogonal projection matrix which want to be obtained. Here, we use greedy iteration algorithm to find $w_k (k = 1, 2, ..., t)$ one by one. $w_k(r)$ stands for the result w_k of the iteration r. In order to remove the absolute value operators in Eq.(3.14), polarity functions $p_l(r)$ and $q_l(r)$ are introduced as

$$p_l(r) = \begin{cases} 1 & \text{if} \quad w_k^T(r)(m_l - m) > 0, \\ -1 & \text{if} \quad w_k^T(r)(m_l - m) \leq 0, \end{cases} \tag{3.15}$$

$$q_l(r) = \begin{cases} 1 & \text{if} \quad w_k^T(r)(x_i^l - m_l) > 0, \\ -1 & \text{if} \quad w_k^T(r)(x_i^l - m_l) \leq 0, \end{cases} \tag{3.16}$$

with the help of $p_l(r)$ and $q_l(r)$, for a special w_k, Eq.(3.14) can be rewritten as

$$\max_{w_k} J_{L_1} = \max_{w_k} \frac{\sum_{l=1}^{C} N_l p_l w_k^T (m_l - m)}{\sum_{l=1}^{C} \sum_{i=1}^{N_l} q_l w_k^T (x_i^l - m_l)}. \tag{3.17}$$

An important property to notice about the objective Eq.(3.17) is that it is invariant with respect to rescaling of the vectors $w_k \to \rho w_k$. Hence, we can always choose w_k such that the denominator is simply $\sum_{l=1}^{C} \sum_{i=1}^{N_l} q_l w_k^T (x_i^l - m_l) = 1$, since it is a scalar itself. For this reason we can transform the problem of objective Eq.(3.17) into the following constrained optimization problem:

$$\begin{aligned} &\max_{w_k} J^*_{L_1} = \max_{w_k} \textstyle\sum_{l=1}^{C} N_l p_l w_k^T (m_l - m), \\ &s.t. \quad \textstyle\sum_{l=1}^{C} \sum_{i=1}^{N_l} q_l w_k^T (x_i^l - m_l) = 1, \\ &\qquad w_k^T w_k = 1. \end{aligned} \tag{3.18}$$

Let

$$f(w_k(r)) = \textstyle\sum_{l=1}^{C} N_l p_l(r) w_k(r)^T (m_l - m). \tag{3.19}$$

First, we want to prove if $w_k(r+1) = arg\max_{w_k} f(w_k)$ then $f(w_k(r+1)) \geq f(w_k(r))$. According to Eq.(3.15), we can get $p_l(r+1)w_k^T(r+1)(m_l - m) \geq p_l(r)w_k^T(r+1)(m_l - m)$, then

$$\begin{aligned} f(w_k(r+1)) &= \textstyle\sum_{l=1}^{C} N_l p_l(r+1) w_k^T(r+1)(m_l - m) \\ &\geq \textstyle\sum_{l=1}^{C} N_l p_l(r) w_k^T(r+1)(m_l - m), \end{aligned} \tag{3.20}$$

since $w_k(r+1) = arg\max_{w_k} f(w_k)$, then

$$\begin{aligned} f(w_k(r+1)) &\geq \textstyle\sum_{l=1}^{C} N_l p_l(r) w_k^T(r)(m_l - m) \\ &= f(w_k(r)). \end{aligned} \tag{3.21}$$

Thus, what we want to do in the next step is to find $w_k(r+1) = arg\max_{w_k} f(w_k)$ where $w_k(r+1)$ stands for the result w_k of the iteration $r+1$. Inspired by the solution of LDA-L2, this problem can be converted into

$$\begin{aligned} &\max_{w_k} f(w_k) = \textstyle\sum_{l=1}^{C} N_l p_l w_k^T (m_l - m), \\ &s.t. \quad \textstyle\sum_{l=1}^{C} \sum_{i=1}^{N_l} q_l w_k^T (x_i^l - m_l) = 1, \\ &\qquad w_k^T w_k = 1. \end{aligned} \tag{3.22}$$

Denote $\vartheta = \sum_{l=1}^{C} N_l p_l (m_l - m)$ and $\phi = \sum_{l=1}^{C} \sum_{i=1}^{N_l} q_l (x_i^l - m_l)$. Consequently, based on augmented Lagrangian method, we can have the following optimization problem:

$$\max_{w_k} \mathcal{L} = w_k^T \vartheta - \lambda(w_k^T \phi - 1) - \tfrac{1}{2}\gamma(w_k^T w_k - 1). \tag{3.23}$$

Based on the Karush-Kuhn-Tucker (KKT) conditions to optimize the question, we can obtain the solution

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial w_k} = \vartheta - \lambda\phi - \gamma w_k = 0, \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = w_k^T \phi - 1 = 0, \\ \dfrac{\partial \mathcal{L}}{\partial \gamma} = w_k^T w_k - 1 = 0. \end{cases} \tag{3.24}$$

Then based on Eq.(3.24), we can obtain the solution of λ, γ and w_k as follows:

$$\begin{cases} \lambda = \dfrac{\vartheta^T \phi}{\phi^T \phi}, \\ \gamma = (\vartheta^T - \lambda\phi^T)\phi, \\ w_k = \dfrac{1}{\gamma}(\vartheta - \lambda\phi). \end{cases} \tag{3.25}$$

Since $p_l(r) \leq 1$ and each element of $w_k(r)$ is less equal to 1, then

$$\begin{aligned} f(w_k(r)) &= \textstyle\sum_{l=1}^{C} N_l p_l(r) w_k(r)^T (m_l - m) \\ &\leq \textstyle\sum_{l=1}^{C} \sum_{i=1}^{d} N_l |(m_l)_i - m_i|. \end{aligned} \tag{3.26}$$

Thus, apparently, $f(.)$ function has an upper bound and increases monotonically. Then, by iteration we can get the optimal w_k in Eq.(3.17). So far, the previous equations can be used to compute the principal vector w_k of $W \in \mathbb{R}^{d \times t}$. And in order to calculate the following vector $w_{k+1}(k = 1, 2, ..., t-1)$, x_i^l should be updated as

$$(x_i^l)^{k+1} = (x_i^l)^k - w_k w_k^T (x_i^l)^k, \tag{3.27}$$

and then be followed by Eq.(3.17), for $k = 1$, we have $(x_i^l)^1 = x_i^l$. This kind of updating rule guarantees that w_{k+1} is orthogonal to w_k and then W is an orthogonal matrix. To justify $w_k^T w_{k+1} = 0$, first we prove that $w_k^T (x_i^l)^{k+1} = 0$ holds. Multiply w_k^T on both sides of Eq.(3.27), then

$$w_k^T (x_i^l)^{k+1} = w_k^T (x_i^l)^k - w_k^T w_k w_k^T (x_i^l)^k, \tag{3.28}$$

since $w_k^T w_k = 1$, then we can have

$$w_k^T (x_i^l)^{k+1} = w_k^T (x_i^l)^k - w_k^T (x_i^l)^k = 0. \tag{3.29}$$

On the other side, w_{k+1} can be represented by $(x_i^l)^{k+1}$, such as

$$w_{k+1} = \sum_{l=1}^{C} \sum_{i=1}^{N_l} \nu_{il} (x_i^l)^{k+1}, \tag{3.30}$$

where ν_{il} are coefficients. Then we can have $w_k^T w_{k+1} = 0$.

Finally, we can get the algorithm of Eq.(3.14) as Algorithm 1.

Algorithm 1 LDA-L1

Require: $X = \{\{x_i^l\}_{i=1}^{N_l}\}_{l=1}^{C} \in \mathbb{R}^{d \times n}, t \leq d$.
  Initialization: $W = [w_1, w_2, ..., w_t] \in \mathbb{R}^{d \times t}, W^T W = I, X_1 = X$.
  for $k = 1 \rightarrow t$ **do**
    $w_k(1) = w_k, r = 1$.
    while not converge **do**
      1. Calculate $p_l(r)$ and $q_l(r)$ according to Eq.(3.15) and Eq.(3.16).
      2. Update w_k according to Eq.(3.25).
      3. r++.
    end while
    Update $X_{k+1} = X_k - w_k w_k^T X_k$.
  end for
  return $W \in \mathbb{R}^{d \times t}$.

It should be noted that there is a possible to obtain a local maximum solution, which is not a global true solution in this iterative procedure. However, since we can set the initial vector w_1 arbitrarily, various initial vectors can be used appropriately to run the LDA-L1 process in several times in order to get the output result which can give the maximum L1 dispersion.

3.6 Evaluation

In this section, our proposed methods will be evaluated on the general database which is described in the previous chapter and some toy set and reconstruction problems are also considered.

In the first evaluation, the total recognition performance of our proposed algorithms in this chapter and some famous feature extraction methods is compared. Two samples of every category are randomly chosen as gallery (training images), and the left ones are applied for probe (testing images). 3 times are performed to randomly select the training group and average recognition precision is recorded.

The final results are concluded in Table 3.1. Here, "LCBP" means that the circle is used to encode the neighborhood instead of multi-scans. "LLCP + LDA-L1" means LDA-L1 is applied to select more compact and discriminative features from LLCP. The approaches "Li et al." and "Pang et al." used L1-norm based 2DPCA and rotation invariant L1 norm based LDA to select LGBP feature, respectively. From this table, we can recognize that our proposed local patterns in this chapter can deal with variations of scale, rotation, illumination more robust than traditional approaches and our proposed algorithms in the previous chapter. In the following subsections, the evaluation of each special variation will also be estimated. One original image is treated as gallery and the corresponding variation ones are used for probe.

Table 3.1: Total accuracy on the whole database.

Method	Accuracy (%)
LBP [6]	56.8
SIFT [3]	69.5
SURF [4]	72.7
ALBP [7]	83.1
LDP [9]	88.5
LGBP [10]	89.3
Li et al. [23]	93.2
Pang et al. [22]	95.1
1DLPMS-B	85.3
LCBP	91.3
LCBPMS	95.7
LLCP	96.2
LLCP + LDA-L1	97.3

3.6.1 Scale

The second experiment shows the performance of scale variation. Table 3.2 shows final recognition rate among several approaches. From this table, we can illustrate that our proposed approaches are comparable or a little better compared to some famous scale invariant features, such as SIFT, SURF and Gabor based feature descriptors.

Table 3.2: Accuracy on the sub-database with variation of scale.

Method	Accuracy (%)
LBP [6]	89.7
SIFT [3]	92.5
SURF [4]	90.7
ALBP [7]	89.5
LDP [9]	92.3
LGBP [10]	93.1
Li et al. [23]	93.9
Pang et al. [22]	94.8
1DLPMS-B	90.3
LCBP	93.3
LCBPMS	94.7
LLCP	95.1
LLCP + LDA-L1	95.3

3.6.2 Rotation

The third experiment shows the influence of rotation on the related methods. Table 3.3 shows final recognition rate among these approaches. From this table, we can illustrate that our proposed approaches are also comparable and better than famous rotation invariant features, such as SIFT, SURF, ALBP and LGBP. While LBP is very sensitive to the rotation changes and LCBP is also a little sensitive to rotation changes.

Table 3.3: Accuracy on the sub-database with variation of rotation.

Method	Accuracy (%)
LBP [6]	13.7
SIFT [3]	87.9
SURF [4]	85.3
ALBP [7]	82.4
LDP [9]	88.5
LGBP [10]	89.8
Li et al. [23]	94.8
Pang et al. [22]	95.9
1DLPMS-B	89.3
LCBP	82.2
LCBPMS	95.2
LLCP	95.3
LLCP + LDA-L1	97.9

3.6.3 Illumination

This fourth experiment shows the illumination effects of the methods. Table 3.4 shows final recognition rate. From this table, we can illustrate that our proposed approaches are much better than some famous algorithms. Since LBP, SIFT and SURF are robust to monotonic lighting variety but very sensitive to non-monotonic illumination changes. The changes of Curvelet filtered images are much more smooth and stable than the original images.

3.6.4 Outliers

In order to show our proposed feature projection by LDA-L1 is invariant to outliers problem, in this subsection, one simple but effective toy set shown in Fig.3.11 is estimated. There are two classes and one outlier in this toy set. Each class has 5 data sample. In the training step, the outlier is treated as class 1. The final results by some traditional methods and our proposed feature projection approach are illustrated in Fig.3.12. From this figure, we can see that the result obtained by our proposed method can separate the class 1, class 2 and the outlier much easier than

Table 3.4: Accuracy on the sub-database with variation of illumination.

Method	Accuracy (%)
LBP [6]	47.7
SIFT [3]	37.5
SURF [4]	45.2
ALBP [7]	57.1
LDP [9]	83.8
LGBP [10]	89.6
Li et al. [23]	92.3
Pang et al. [22]	93.7
1DLPMS-B	81.3
LCBP	90.3
LCBPMS	92.3
LLCP	94.1
LLCP + LDA-L1	97.6

the results by traditional methods. That means, our proposed feature projection are more effectiveness to solve outliers problem compared to some famous approaches.

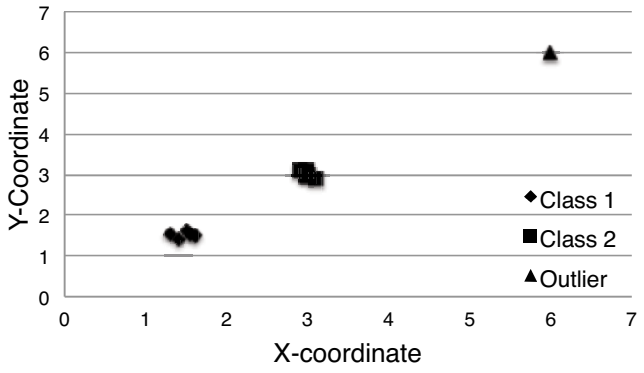


Fig. 3.11: Toy Set.

3.6.5 Occlusion

In order to evaluate the occlusion invariant property of our proposed feature projection method, one reconstruction problem is considered as shown in Fig.3.13. From

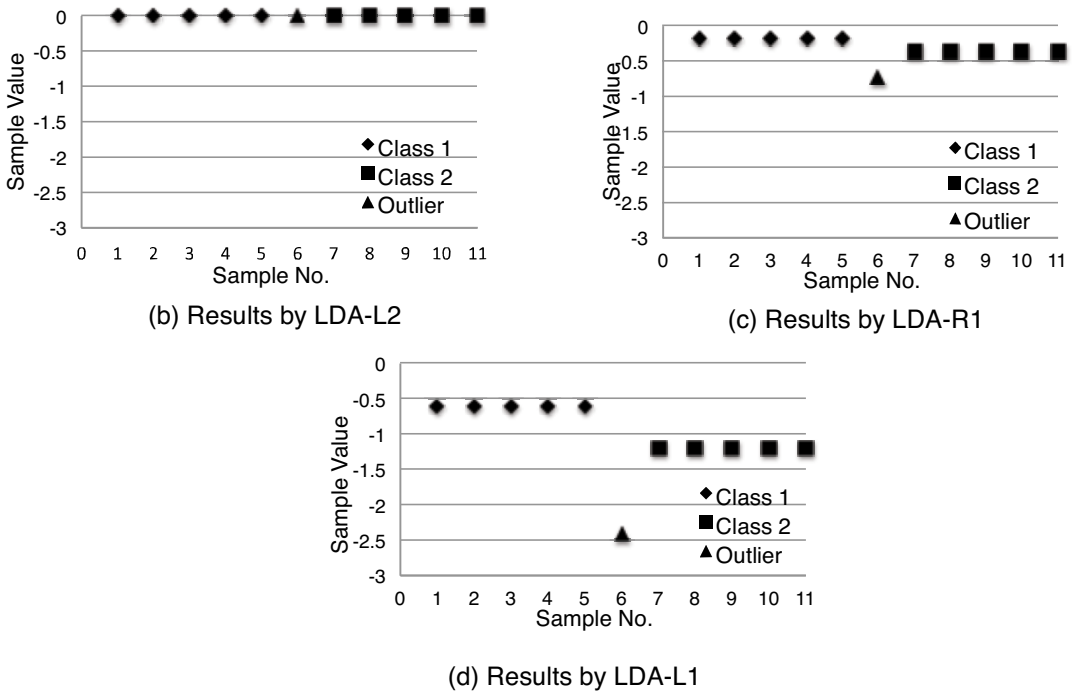


Fig. 3.12: Projection results of toy set.

this figure, we can see that the reconstructed image by LDA-L1 almost has no occlusion problem after low rank projection compared to the original occluded image. Note that in some sense, the occlusion can be treated as a special case of the outliers.

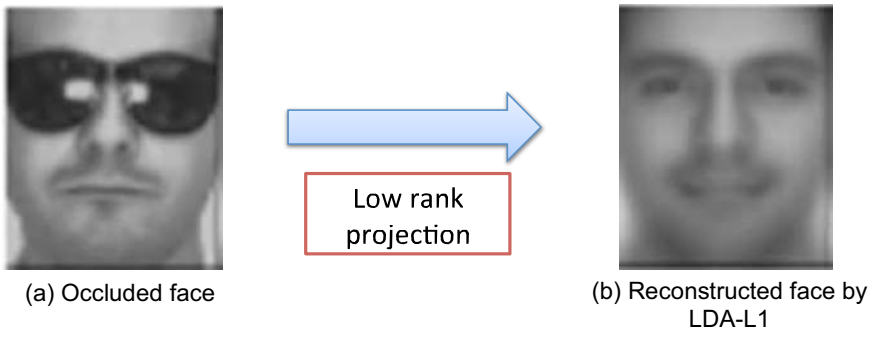


Fig. 3.13: Reconstructed image by LDA-L1 under occlusion condition.

3.7 Conclusions

In this chapter, firstly, the previous multi-scans based feature descriptor in chapter 2 is extended into Curvelet frequency space. The Curvelet and multi-scans based local patterns are mutually complementary because multi-scans based local patterns can capture the local appearance detail, whereas Curvelet can extract global information. Two novel approaches called Local Curvelet Binary Patterns by Multi-Scans(LCBPMS) and Learned Local Curvelet Patterns (LLCP) for presenting the local feature are proposed. The Curvelet represented images can better generate the curved edge and hyperplane singularities than some traditional methods. LCBPMS uses some predefined patterns to encode the image while LLCP learns some codebooks from sampled patches that are regraded as pattern-specific and more desirable for pattern perception tasks. During image representation phase, multi-mapping strategy is used in LLCP, which is more reasonable than traditional one to one mapping. Secondly, one novel feature projection method of LDA based on L1-norm optimization, which can deal with the outliers and occlusion problem as well as select more compact and discriminative feature from high dimension, is proposed. LDA-L1 is aimed to get projections that maximize the L1-criteria at the projected feature dimension instead of the traditional L2-norm to better characterize the between-class separability and within-class compactness. The proposed L1-norm optimization technique avoids to compute the eigen-decomposition problem and is easy to implement. In more specification, it first suppresses the negative effects of outliers or occlusion, second it has no 3S problem and third it is invariant to rotations. Experiments on some databases show that our proposed methods have better performance than relevant ones.

4

Face Recognition with Multi-Scans based Invariant Feature Descriptors

4.1 Introduction

Over the past twenty years or so, face recognition has been a hottest field in the research of pattern recognition. Compared with some other biometrics [40], for instance iris or fingerprint, face recognition has great advantage in high-universality, high-collectability, high-acceptability, and low-circumvention. Due to these advantages of face recognition, many researchers have been developed some powerful systems to assist and help our human beings in the areas of image analysis, computer vision, la . Although face recognition system has shown impressive advantage in literature and industry, face recognition issue is still considered to be unsolved and an open and challenge topic due to the outside factors changes under various conditions, such as illumination, facial expression, noise, pose, occlusion and so on [1] [41]. Generally, based on how to represent the face elements, current researches can be briefly separated into two groups: global feature-based and local feature-based approaches. In detail, the element in the global

feature-based method is related to the whole input facial picture and can represent the whole face, while the element in the local feature-based method is obtained from some local area in the facial image. However, it is should be noted that there is no clear separation between these two categories. For example, the local feature-based method can be treated as the global method when the size of the area is increased to the whole image. Most of the global methods are subspace methods that reduce the dimension of the image. The Principle Component Analysis (PCA) [14], Independent Component Analysis (ICA) [42], Linear Discriminant Analysis (LDA) [16] and Discrete Cosine Transform (DCT) [43] have been the famous method in this class. PCA uses an eigenvalue subspace to project the whole image into several weights, and employs the distances between these weights to recognize faces, while the higher order statistics is taken into account in ICA that is suitable to learn complex structure on the database. LDA considers the difference both between-class and within-class matrix, and DCT could remain more linear property. Recently, in order to decrease time complexity and obtain two-dimension structure information of images, 2DPCA [44], 2DLDA [45], and 2DPCA-L1 [22] were proposed. However, all of these methods are global representation of images that are sensitive to global changes of images, such as, illumination and expression.

In order to overcome these drawbacks of global feature based methods, local matching approaches are presented in face recognition [6], [46] and other visual recognition tasks [47] that are invariant to illumination and expression issues. The basic concept of local matching approach is that it is first to extract some features, and secondly to category the samples based on the measurement of the corresponding local statistics. The most famous method in this field is Local Binary Patterns (LBP) [6]. Recently, Zhang [9] proposed Local Derivative Pattern(LDP) that is a common framework for encoding micro-pattern feature vectors that considers local derivative variations. LDP could obtain more effective structure information com-

pared to LBP that only uses the first order local pattern. However, all of these patterns use circle to encode the local neighborhood pixels, that means, just one kind of spatial information around is considered. Specially for LDP, in previous steps, local derivative variations is used to capture more detailed information, however it has high time complexity. From the reconstructed images in ref. [9] by LDP, more detailed information also took more noise into account and key parts of our faces could not separate clearly. At final step, also circle encoding was applied. Another problem of the existing approached in the field is that if the number of the neighborhood pixels enlarged a lot, which makes the number of patterns be huge, then the encoding and calculation of these patterns are time consuming. For example, LBP [6] locates a block into each pixel in the whole sample through thresholding a 3×3 neighborhood at every point with the center one, as a result a binary pattern can be got in final. Generally, the symbol (P, R) is used in this encoding step where P denotes the number of sampling points while R is the circle radius used for sampling [6]. If P is small, then we can not get discriminant and spatial information in this block as much as possible, while P is large, encoding seems difficult and lots of histogram of the patterns are equal or similar to zero which can be noted as noise and are not useful for representation. As an example, if $P = 16$. then $2^{16} = 65536$ patterns will be generated. When the resolution of an image is 100×100, the average magnitude of the histogram is just $10000/65536 = 0.15$. Under this condition, many patterns will be useless.

To address these issues, in chapter 2, based on multi-scans encoding model, a novel, theoretically and computationally simple approach, which is called 1D local patterns by multi-scans(1DLPMS) [48], is proposed in our work. First, multi-scans are used to capture the different spatial as well multi-resolution and multi-orientation information on the image. Compared to the traditional ones [6] which only used circle to encode the neighborhood pixels, multi-scans based feature descriptor can

keep more spatial information and it is invariant to scale, rotation and illumination. Then based on varied encoding rules, some 1D local patterns are proposed to encode local features with less time consuming. To simplify the computation of 1DLPMS , grouped 1D local patterns by multi-scans(G1DLPMS) is proposed, which separates 1DLPMS into several groups and only the co-occurrences of these groups are used.

Recently, since Gabor wavelet has good characteristics in space frequency, space position and direction selectivity, many researchers have been proposed local patterns from Gabor feature to represent facial images, such as Histogram of Gabor Phase Patterns (HGPP) [11], Local Gabor Binary Patterns (LGBP) [10] and Learned Local Gabor Patterns (LLGP) [12]. The key different idea between HGPP, LGBP and LLGP is that some pre-defined patterns are used in HGPP, LGBP while some learned patterns are applied in LLGP. No matter which method, in final, the facial image can be encoded into some vectors according to the pre-defined or learned patterns. However, the common issue of these methods is that the feature dimension is very large due to Gabor decomposition, and there is no clear discuss about why the Gabor transform is the best frequency space for representing facial images? Actually, Gabor transform cannot well represent curve singularity of human facial images since Gabor wavelets are very powerful in representing samples with isolated point singularities, but they are failed to represent line or curve singularities.

In Ref. [29] , Candes et al. proposed Curvelet Transform, which can capture hyperplane and curved edge singularities, to overcome the weakness of Gabor wavelets in high dimensions. Different from Gabor wavelet, the basic representation elements in Curvelet Transform is edge, which is strongly anisotropic. As illustrated in Ref. [29], Curvelet Tranform can represent the curved singularities optimally in the higher dimension. The fine and detail coefficients, which are very powerful to detect curves in images, are strongly orientation-sensitive. In [49], comparison of wavelet, Gabor wavelet and Curvelet transformation to recognize faces under illu-

mination and expression changes was discussed and concluded that Curvelet was a better choice, since compared to wavelet, the Curvelet transform has the property to represent the images more sparsely. And the most advantage of Curvelet is that it could extract high degree of orientation, anisotropy and time frequency resolution, which can generate more powerful and effective descriptors in images or textures.

Generally, in literature, Curvelet feature based face recognition methods are just applying some subspace to project the Curvelet coefficients. For example, in [50], [51] and [52], PCA was used while in [53], LDA was considered. None of these methods is face-specific, and does not consider the spatial information about our face. In addition, these methods are sensitive to illumination and facial expression change.

In chapter 3, unlike the common methods which used Gabor wavelet to transform facial images into frequency space and overcome the problems of traditional Wavelet and Curvelet feature based face recognition, Local Curvelet Patterns are studied in our work. Based on Curvelet filtered facial images, two local Curvelet patterns called Local Curvelet Binary Patterns by Multi-Scans (LCBPMS) and Learned Local Curvelet Patterns (LLCP) are proposed. Next, the facial image can be encoded into multi-pattern maps according to 1DLPMS operator and some learned patterns, respectively. In final, the histograms of the patterns for all the regions are concatenated together to construct a histogram sequence to represent the input facial image. During face representation part, multi-patterns are used to encode the input patch that is sampled from Curvelet filtered facial image in LLCP, since these encoded multi-patterns have close similarities with the input patch.

In this chapter, since compared to vector, matrix is more suitable to present our face to keep the spatial structure of our facial image; our proposed LDA-L1 in chapter 3 is first extended from 1D into 2D. Next, in order to get robust classification result, one effective classifier called Weighted Histogram Spatially constrained Earth Mover's Distance (WHSEMD) which illustrates the discriminative powers of various

images regions, the different patterns and the different spatial information of images, is proposed.

To evaluate our proposed methods, face identification, gender estimation and facial expression recognition are studied. Experimental results on face recognition based on six famous and challenging face databases - ORL [54], FERET [1], LFW [55], FRGC [56], AR [2] and JAFFE [57] show that our proposed methods outperform some state-of-the-art algorithms.

The remainder of this chapter is organized as follows: the variants of L1-norm based LDA will be described in section 2 and section 3. Our proposed classifier called WHSEMD will be introduced in section 4, Experiments about face identification, gender estimation and facial expression recognition will be reported in section 5, section 6 and section 7, respectively. Finally, conclusions are discussed in section 8.

4.2 L1-norm based Two Dimensional Linear Discriminant Analysis

L1-norm based Two Dimensional Linear Discriminant Analysis (2DLDA-L1) aims to find two optimal transformation matrices $W_1 = [w_1^1, w_2^1, ..., w_{t1}^1] \in \mathbb{R}^{w \times t_1}$ and $W_2 = [w_1^2, w_2^2, ..., w_{t2}^2] \in \mathbb{R}^{h \times t_2}$ to maximize the following objective function

$$\begin{aligned} \max_{W_1,W_2} J_{L_1}^{2D} &= \max_{W_1,W_2} \frac{\sum_{l=1}^{C} N_l ||W_1^T (m_l^{2D} - m^{2D}) W_2||_{L1}}{\sum_{l=1}^{C} \sum_{i=1}^{N_l} ||W_1^T ((x_i^l)^{2D} - m_l^{2D}) W_2||_{L1}} \\ &= \max_{W_1,W_2} \frac{\sum_{l=1}^{C} N_l |W_1^T (m_l^{2D} - m^{2D}) W_2|}{\sum_{l=1}^{C} \sum_{i=1}^{N_l} |W_1^T ((x_i^l)^{2D} - m_l^{2D}) W_2|}. \end{aligned} \tag{4.1}$$

In 2DLDA-L1, the sample is treated as a matrix instead of the vector in LDA-L1. $(x_i^l)^{2D} \in \mathbb{R}^{w \times h}$, where w and h denote the width and height of samples, respectively. $m_l^{2D} = (1/N_l) \sum_{i=1}^{N_l} (x_i^l)^{2D}$ is the average of the samples which is assigned to the l-th class, and $m^{2D} = (1/n) \sum_{l=1}^{C} N_l m_l^{2D}$ is the global mean value of the all samples. The key problem to solve this objective function is that it is not convex. But fortunately, when one parameter is fixed, the objective function of the other parameter is convex.

Then based on the solution of LDA-L1, projection matrix W_1 and W_2 can be solved one by one. More specifically, firstly, we can fix W_2 to solve W_1 and then fix W_1 to solve W_2.

4.3 L1-norm based Block Linear Discriminant Analysis

In this section, the L1-norm based optimization method is used to Block LDA, which is called L1-norm based Block Linear Discriminant Analysis (BLDA-L1). Instead of row separation in 2DLDA-L1, the basic unit - block is applied in BLDA-L1, since the block has strong correlation in the image compare to row. Thus, in BLDA-L1, the image is firstly separated into some blocks, and then converts the block elements into row vectors as same used in 2DLDA-L1.

Fig.4.1 gives an example. Assume that we have original 6 by 6 image as Fig.4.1(a) (the number is denoted the location of each pixel), and the block size is set 3 by 3. Then we can non-overlapping divide the original image into 4 blocks and convert it into 9 by 4 image as Fig.4.1(b). Each row in Fig.4.1(b) stands for one block in Fig.4.1(a). That is, BLDA-L1 can be converted into 2DLDA-L1, and then we can use the algorithm for 2DLDA-L1 to solve it.

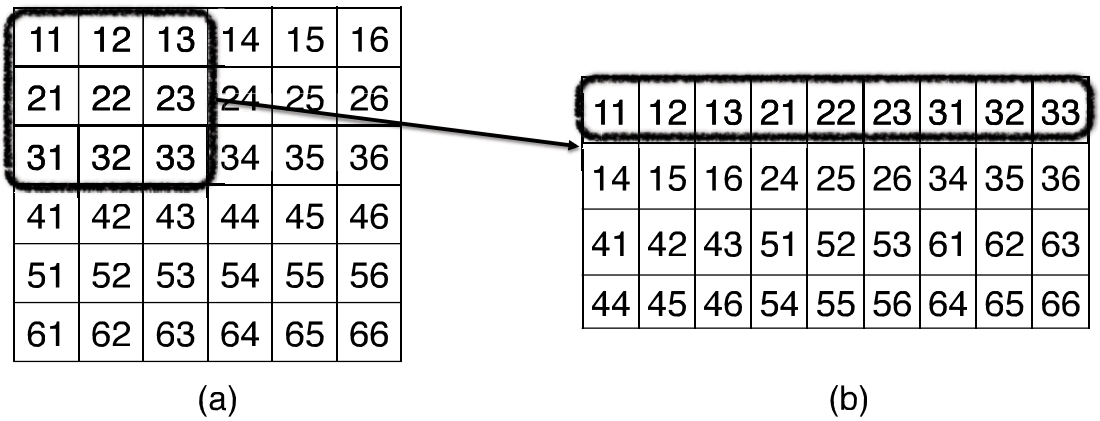


Fig. 4.1: Sample of BLDA-L1 (Block size is 3 by 3).

Some points should be noted. If the block is selected as each row of the original

image, then BLDA-L1 is same as 2DLDA-L1. If the whole image is treated as one block, then BLDA-L1 is same as LDA-L1. Thus, BLDA-L1 can be considered as a general framework of LDA-L1 and 2DLDA-L1.

4.4 Weighted Histogram Spatially constrained Earth Mover's Distance

In this section, a new classifier called Weighted Histogram Spatially constrained Earth Mover's Distance (WHSEMD) is proposed. It is an improvement version of Spatially constrained Earth Mover's Distance (SEMD) [58]. To reduce the time complexity, probe Image I_P and gallery Image I_G are presented with non overlapping and overlapping regions, respectively, at the same time, in order to decrease the time complexity, only a small local searching neighborhood in galley I_G is considered for each region in probe I_P. Thus, with these two constraints, the parameters of our approach are greatly reduced and the time cost is significantly decreased.

The (directional) measurement from probe I_P to gallery I_G can be calculated by

$$d(I_P \rightarrow I_G) = \sum_a \sum_i \sum_{j \in N^r(i)} w_{a,i} f_{ij} \check{g}\left(H_{a,i}^{I_P}, H_{a,j}^{I_G}\right), \tag{4.2}$$

where $w_{a,i}$ is the weighing value for different facial part corresponding to layer a and region i, f_{ij} is the weighting value based on the histogram $H_{a,i}^{I_P}$ and histogram $H_{a,j}^{I_G}$, $N^r(i)$ is the index group of r nearest regions based on region i in I_P in the spatial dimension. $\check{g}(.)$ is the measurement between two histogram defined as

$$\check{g}\left(H^{I_P}, H^{I_G}\right) = \sum_{k=0}^{K-1} \bar{\omega}_k min\left(H^{I_P}(k), H^{I_G}(k)\right), \tag{4.3}$$

here, $\bar{\omega}_k$ is the weighing value corresponding to pattern k. The calculation of $w_{a,i}$ and $\bar{\omega}_k$ is presented in [10] and [12], respectively. Note: in face representation with

LCBPMS, $\bar{\omega}_k$ is always set to 1.

Generally, for every region in image I_P, nearest regions in the spatial dimesion in image I_G, bounded by a distance of $\sqrt{2}$ can be got in this work. Low pass filter (Gaussian Kernel is used in our study) is applied around to reduce noise and get a smoother score for classification. Thus, WHSEMD is slightly expect to be invariant to pose change.

4.5 Face Identification

In this section, our proposed multi-scans based invariant feature descriptors and classification method are evaluated on face Identification. F l-known databases - ORL [54], FERET [1], FRGC [56] and AR [2] are considered.

4.5.1 ORL Database

The ORL database consists of facial images of 40 different people and each individual has 10 different images. Thus, totally 400 images are used in this study with some variations in each subject category. Firstly, the faces of each category are captured in different times and slightly changes in illumination. Secondly, various facial expressions are recorded, such as whether smiling or not, closed eyes or not, with glass or not. Thirdly, different poses with a tolerance for raw, pitch and roll of up to 20 degrees are also considered. Fourthly, the scale sizes are changed to up to about 10 percent. The total facial images are gray-level and normalized into 112×92 resolution. Some original examples are shown in Fig.4.2.

Each image is divided into 16 regions. Some reconstructed images by our proposed feature descriptors in chapter 2 are list in Fig.4.3. Compared to LBP, combined image have clearer edge and less noise, especially in 1DLTP and 1DLQP.

In the first experiment, the leave-one-out strategy is used. The result is listed in Table 4.1. The notation "Num." means the number of patterns and "time" corre-

Fig. 4.2: Samples from ORL database.

sponding to matching time. It can be seen that -T and -Q achieve higher accuracy than LTP with less patterns and computational complexity. And the best one is G1DLPMS-T with the trade-off between accuracy and computational complexity. Same parameters P, $P^{'}$ and $P_G^{'}$ will be used in the following.

Table 4.1: Leave-One-Out on ORL database.

Method	Accuracy(%)	$P/P^{'}$	$P_G^{'}$	Num.	time(ms)
LBP [6]	97.3	8	/	256	109
1DLPMS-B	98.5	4	/	16	7
G1DLPMS-B	97.8	6	2	32	13
LTP [26]	98.8	8	/	6561	2562
LDP [9]	98.2	32	/	1024	412
1DLPMS-T	99.0	4	/	81	39
G1DLPMS-T	99.5	8	4	162	81
1DLPMS-Q	99.3	4	/	256	123
G1DLPMS-Q	99.5	6	4	272	137

In second experiment, some $D_{num}(D_{num} = 1, 2, 3, 4, 5)$ image(s) of each person will be randomly selected to learn, while the left images for testing. To compare our method with LBP and other famous approaches, five tests are evaluated with a different number of training set and mean rate is recorded. Table 4.2 shows the

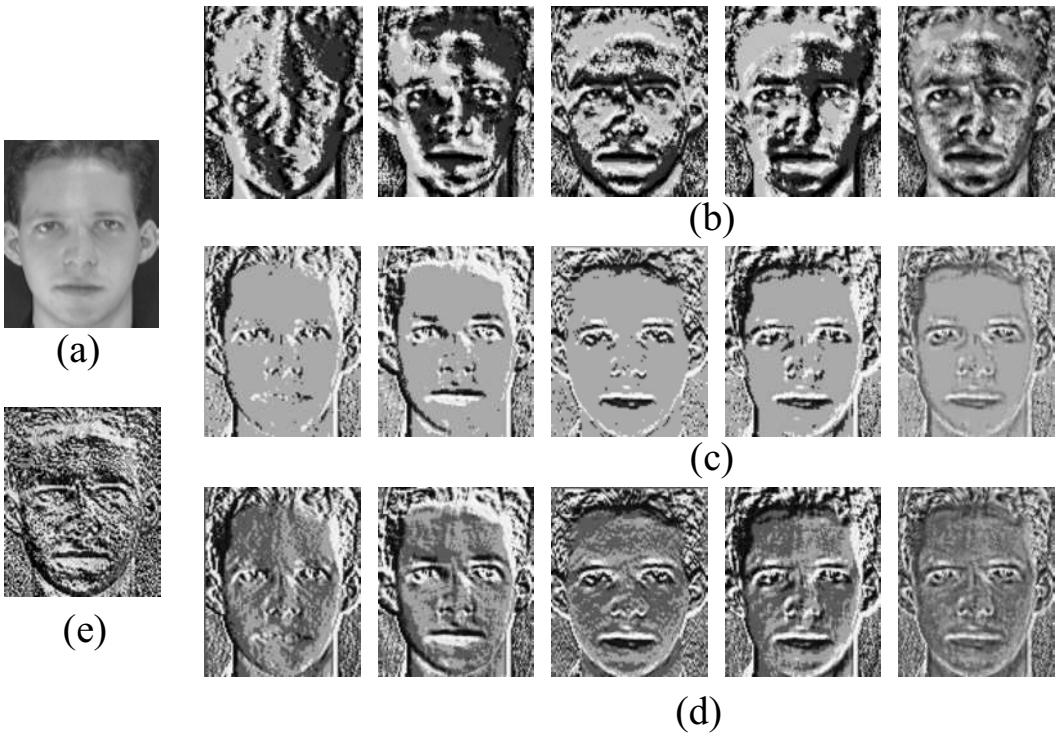


Fig. 4.3: Reconstructed images. (a)Original image. (b),(c),(d),(e) are reconstructed image(s) by 1DLBP, 1DLTP ,1DLQP and LBP respectively (scan order from left to right in each row is S1, S2, S3, S4, the final image in each row is the combined image using these 4 scans).

accuracy($\tau = 2$ in this case). We can see that our proposed methods achieve better performance. Among our proposed methods, G1DLPMS-T and G1DLPMS-Q are outstanding since they can keep more discriminable information than others.

In third study, we try to find which scan order can keep key significant spatial information. In this case, G1DLPMS-T is selected, and $D_{num} = 2$, the recognition rates are 85.9%,87.2%, 92.2%, 81.8% for S1, S2, S3, S4 respectively ($\Delta\theta = 45^o$). S3 is the best since it can catch more discriminant features in our face, such as eyes and mouth. And also, this point can be seen from the reconstructed images. Note that S3 gets more useful spatial information than basic LTP with less number of patterns.

In order to consider how parameter P' and h affect the recognition result, several experiments are carried out as Table 4.3 and Table 4.4. From these tables, larger

Table 4.2: Accuracy (%) on ORL- D_{num} (1,2,3,4,5) for gallery and others for probe.

Method \ D_{num}	1	2	3	4	5
PCA	66.9	84.7	88.2	90.8	93.5
LDA	75.2	85.6	87.5	91.7	92.5
LBP [6]	74.7	82.5	87.8	92.0	94.0
LGBP [10]	77.5	87.8	88.6	92.9	96.0
L	77.5	86.9	90.7	94.6	97.0
ALBP [7]	75.0	82.7	88.6	92.5	94.0
DLBP [28]	73.6	83.1	88.6	92.9	95.0
LDP [9]	78.9	83.1	87.9	93.5	96.0
1DLPMS-B	79.7	83.2	89.0	92.9	95.5
G1DLPMS-B	77.5	85.3	88.5	93.8	96.0
1DLPMS-T	81.9	89.1	91.1	94.6	97.5
G1DLPMS-T	86.7	93.0	93.6	97.1	98.0
1DLPMS-Q	85.0	90.0	92.0	95.0	97.5
G1DLPMS-Q	87.2	91.9	93	97.1	98.0
LCBPMS	86.2	90.9	91.6	96.2	99.0
LLCP	91.2	93.7	94.2	98.9	100.0

P' can get better performance since it can get more local information. In the other hand, larger P' can increase the time complexity. For the sake of the balance of speed and accuracy, $P' = 4$ is used in following. Meanwhile, a little larger h can get better performance since it can capture local information more efficiently.

Table 4.3: Influence of parameter P' by 1DLPMS-B, one image for training.

P'	Accuracy (%)
2	79.7
4	82.2
6	83.9
8	83.9

Since the different parameter τ can catch different proposed patterns, in order to evaluate how it affects the accuracy, Table 4.5 lists some comparisons. (Note: the other parameters are same, and it just uses one image for training,)

From this table, we can see that when τ is around 4.5, it can obtain best accuracy.

Table 4.4: Influence of parameter h by 1DLPMS-B, one image for training.

h	Accuracy (%)
1	80.3
1.5	80.3
2	82.2
2.5	83.3
3	83.9
3.5	83.6
4	83.5

Table 4.5: Influence of parameter τ by 1DLPMS-T, one image for training.

τ	Accuracy (%)
1.5	80.6
2	82.5
2.5	83.1
3	84.7
3.5	84.2
4	85.0
4.5	85.3
5	84.4
5.5	83.9
6	83.6

If τ is small, that means, we will set the close area very large, but in fact, the surrounding intensity in the facial image is closed. Thus, in this case, we can not discriminant the close area and non-close area clearly. Otherwise, if τ is very large, that means the close area is very narrow, also, we can not discriminant the close area and non-close area clearly. Thus, in our experiments, the parameter τ is around 4.5.

Next evaluation will estimate how the parameter ξ affects the final accuracy, since different ξ will get different edge response information. The result is shown in Table 4.6 ($\Delta\theta = 45^o$). From this table, we can see that when $\xi = 4$, the performance is almost stable and enough orientations can be got. Then in the following experiments, ξ is set to 4 by default.

The next experiment over the ORL database [54] is to compare the feature pro-

Table 4.6: Influence of parameter ξ by 1DLPMS-T, one image for training.

ξ	Accuracy (%)
1	82.7
2	82.9
3	84.9
4	85.3
5	85.3
6	85.1
7	85.7
8	85.9

jection performances and reconstruction errors of LDA-L2, LDA-R1 and our proposed methods. All the samples are gray-level and normalized to 32×32 resolution. Among these 400 images, 30 percent are randomly chosen and occluded by a rectangular noise which is consist of random white and black dots with the size of 10×10 resolution, located at a random position. In order to give a better description, some learning samples are shown in Fig.4.4. 3 images per person are used for training and others are for testing. Here, simple 1-nearest-neighbor(1NN) classifier is used for the final classification. The performance is shown in Fig.4.5, where x-axis is associated with the reduced dimension and y-axis is corresponding to the accuracy. The average number of iterations for LDA-L1 is 5.1 while 9.7 for LDA-R1. From this figure, we can see that our proposed method is the outstanding one and can obtain about 10 percent or 40 percent than LDA-R1 and LDA-L2, respectively. Moreover, in this figure, when the reduced dimension is very small, our proposed method can get significant performance. In order to see how the accuracy changes in small dimension, another experiment is carried out and the result is shown in Fig.4.6. From this figure, we can see more clear about the effectiveness of our proposed method.

Table 4.7 and Table 4.8 shows the accuracy with the change of selected feature dimension and block size by 2DLDA-L1 and BLDA-L1, respectively. From these two tables, we can illustrate that the accuracy of two dimensional based LDA is higher

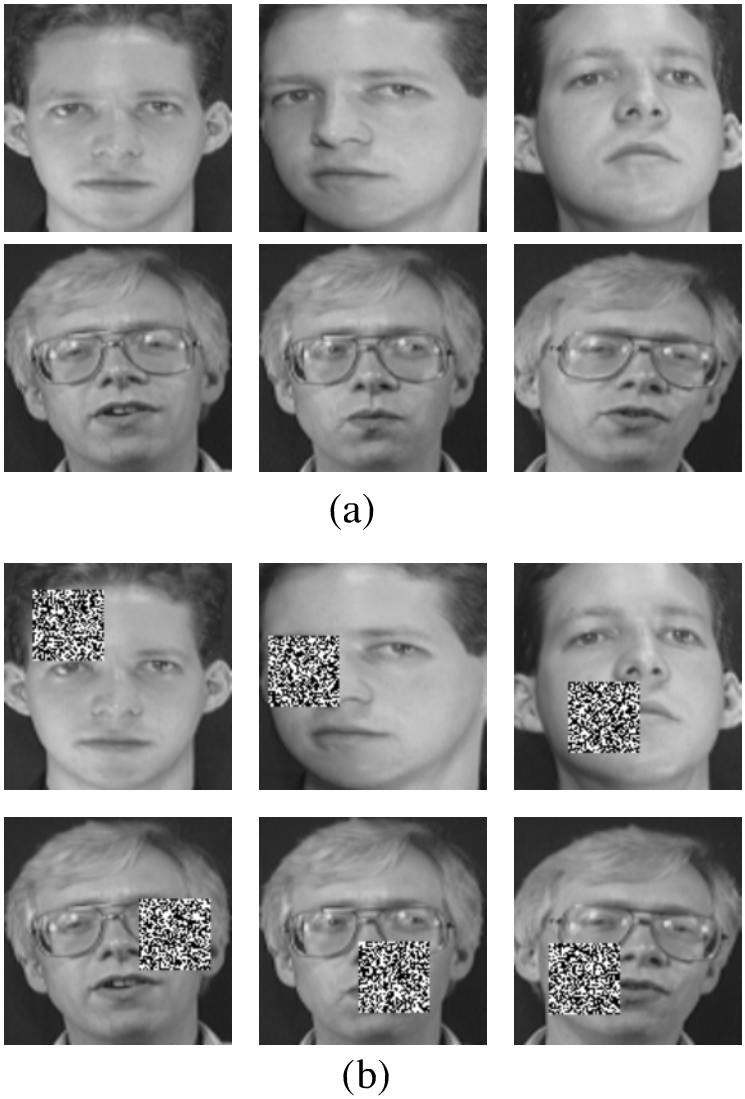


Fig. 4.4: ORL database. (a) Original Images. (b) Corresponding Images with occlusion.

than one dimension based LDA, and in Table 4.8, smaller or larger block size will decrease the recognition rate. The reason may be that if the block size is small, then the correlation within each block is weak, and the correlation between blocks will be weak if large block size is applied. Finally, the accuracy on ORL database is concluded in Table 4.9, and our proposed methods has higher performance than traditional ones.

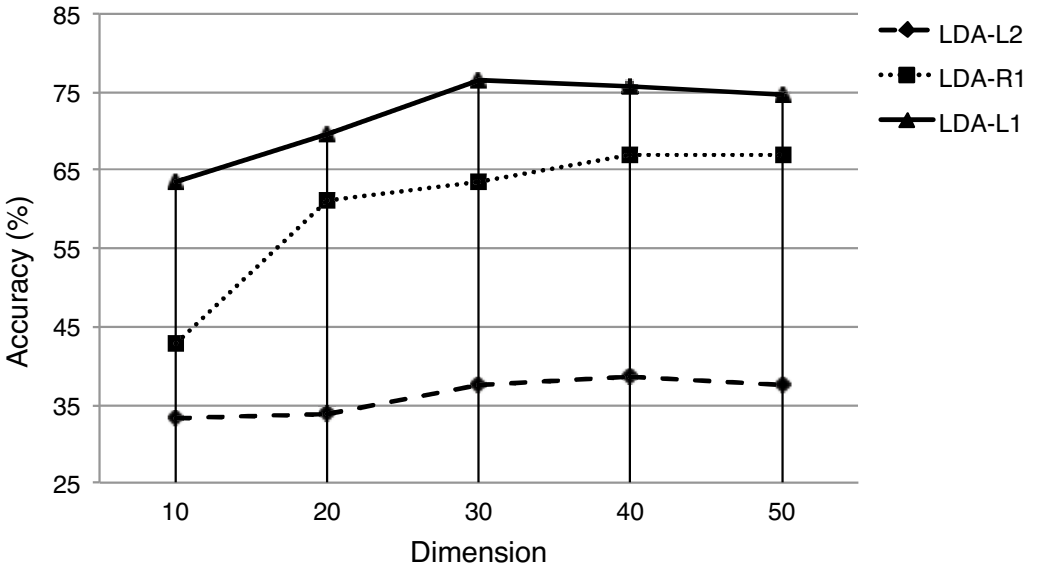


Fig. 4.5: Classification results on occluded ORL database.

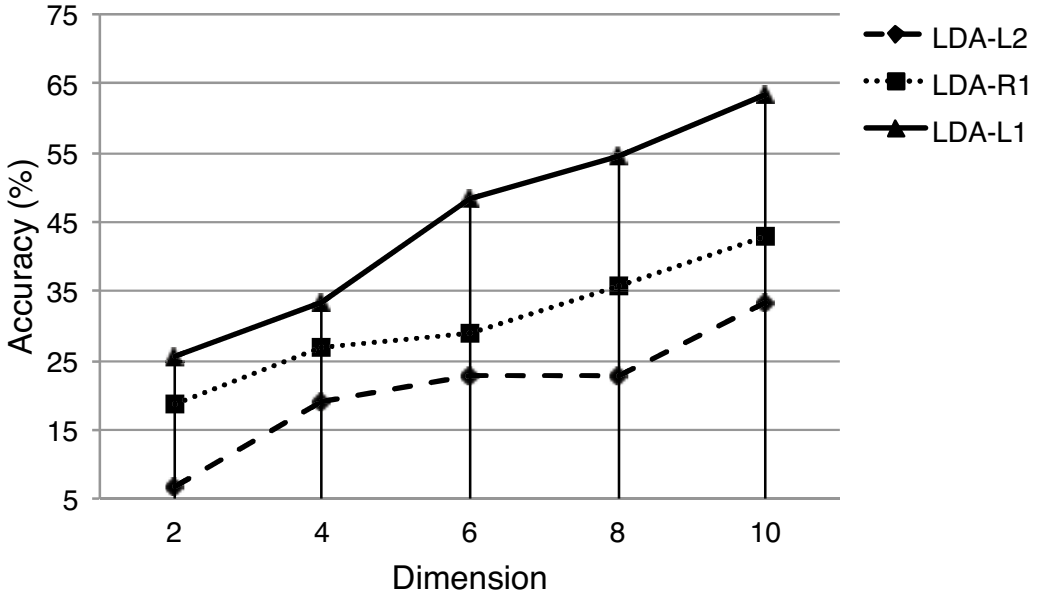


Fig. 4.6: Classification results for small dimension on occluded ORL database.

In order to illustrate the power of our proposed classification approaches in unoccluded case, several evaluations are carried out. Fig.4.7 illustrates the relationship between the accuracy and the reduced dimension. And Table 4.10 shows the compar-

Table 4.7: Classification results with different dimensions on occluded ORL database by 2DLDA-L1 method.

Dimensions	Accuracy (%)
2×2	77.2
3×3	79.8
4×4	80.7
5×5	81.9
6×6	81.2

Table 4.8: Classification results with different block size by BLDA-L1 method on occluded ORL database.

Block Size	Accuracy (%)
2×2	78.2
4×4	83.2
8×8	82.9
16×16	80.9
32×32	76.6

ison among several traditional approaches and our proposed methods. From these experiments, we can generate that our proposed methods are not only effective to deal with occlusion problem, but also powerful in un-occluded case.

In the next experiment, in order to measure how well the face can be represented, face reconstruction issue is considered. The performance of various approaches , such as LDA-L2 [16], LDA-R1 [23], LDA-L1, 2DLDA-L1 and BLDA-L1, are applied and compared. The average reconstruction error is calculated between the original un-occluded faces and the faces reconstructed by the fraction of features as Eq.(4.4) and Eq.(4.5) for one dimension based and two dimension based LDA, respectively,

$$e_1(m) = \frac{1}{n}\sum_{i=1}^{n} ||x_i^{org} - \sum_{k=1}^{t} w_k w_k^T x_i||_2, \tag{4.4}$$

$$e_2(m) = \frac{1}{n}||(X^{2D})^{org} - W_1 W_1^T X^{org} W_2 W_2^T||_2. \tag{4.5}$$

Here, n is defined as the total number of samples with value 400 in this setup, x_i^{org}

Table 4.9: Accuracy on occluded ORL database.

Method	Accuracy (%)
LDA-L2 [16]	38.6
LDA-R1 [23]	67.1
2DPCA-L1 [22]	79.1
LDA-L1	76.6
2DLDA-L1	81.9
BLDA-L1	83.2

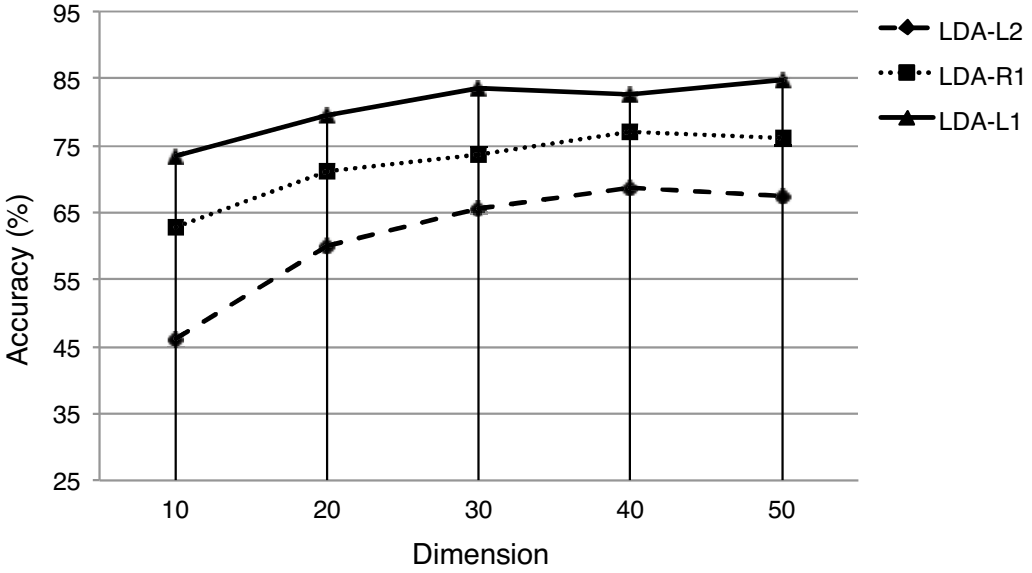


Fig. 4.7: Classification results on un-occluded ORL database.

and x_i are the i-th original un-occluded sample and the i-th sample applied in the training, respectively, the t is defined as the number of extracted features. $(X^{2D})^{org}$ and X^{org} are original un-occluded matrix based image and occluded matrix based image applied for training, respectively. Fig.4.8 illustrats the average reconstruction errors which are computed by various numbers of extracted features. From this figure, we can see clearly that even if the number of extracted features is very small, the average reconstruction errors of our proposed approaches are much smaller than LDA-L2 and LDA-R1 approaches. The difference between our proposed methods and traditional methods are apparent and BLDA-L1 is the most outstanding one.

Table 4.10: Accuracy on un-occluded ORL database.

Method	Accuracy (%)
LDA-L2 [16]	67.5
LDA-R1 [23]	77.1
2DPCA-L1 [22]	85.1
LDA-L1	84.7
2DLDA-L1	89.3
BLDA-L1	91.4

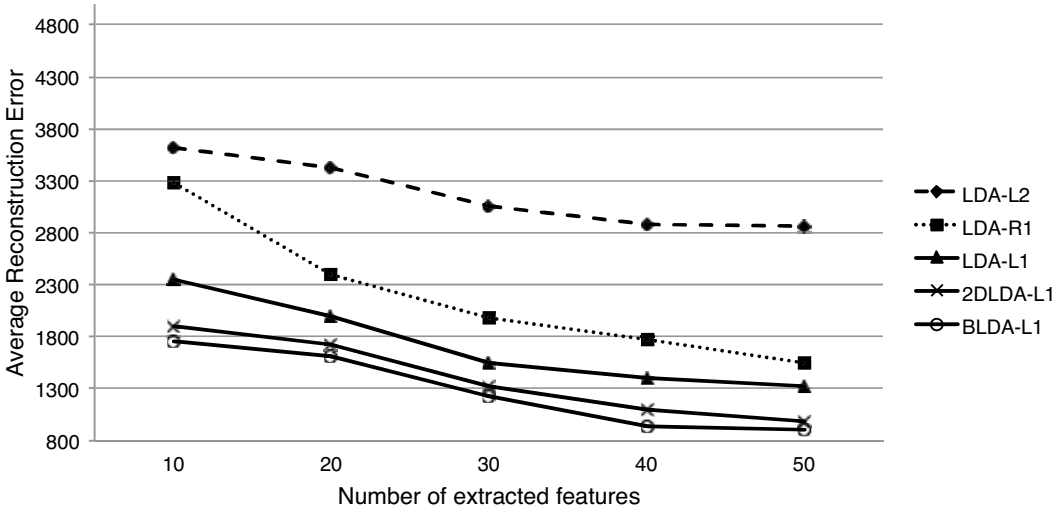


Fig. 4.8: Average reconstruction errors for ORL database.

4.5.2 FERET Database

In our study, regular frontal image called 'Fa' set is used for making galley set, the size of galley in our study is 581. The 'Fb' probe set is for analyzing the effectiveness of a various facial expression on recognition performance. The size of 'Fb' probe set is 580. The size of 'Dup I', which consists of all other frontals of the subjects taken several days or even years later than 'Fa', is 475. The 'Dup II' probe set that is captured at least one year later than 'Fa' is a subset of the 'Dup I' probe set and the size is 118. The images are copped by eyes and mouth but no normalization methods are applied (Note that these cropped images can also be directly detected by some

traditional face detection algorithm). Each face is resized to 100 × 100 and divided into 25 regions. Fig.4.9 gives some samples. (the first two images are from 'Fa' and

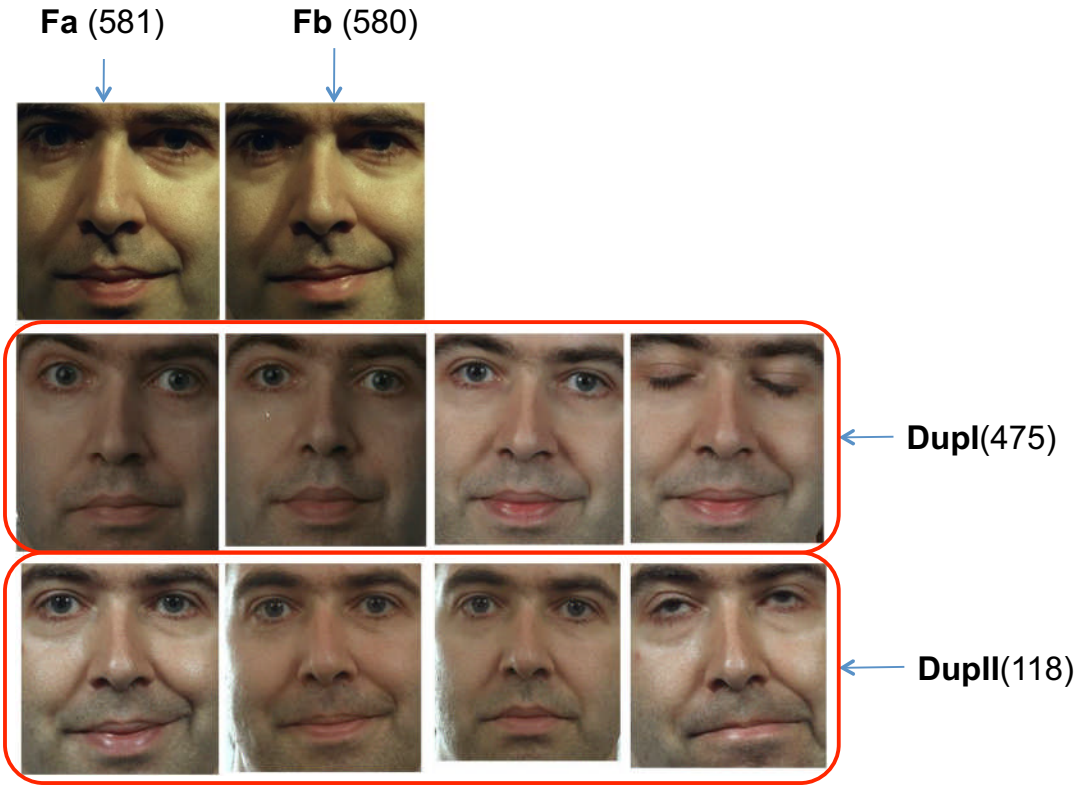


Fig. 4.9: Sample cropped facial images of one individual from FERET [1] database

'Fb' set, respectively. The images in second row are located in 'Dup I' set, while the images in last row come from 'Dup II' probe set.) In this experiment, we just use one image in set 'Fa' for training, and 'Fb', 'Dup I' and 'Dup II' for testing. The result is shown in Table 4.11. G1DLPMS-T and G1DLPMS-Q are two comparable ones while G1DLPMS-T uses less number of patterns. Weighted LCBPMS and LLCP are two outstanding ones while weighted LLCP can get the best performance. Here the "weighted" means WHSEMD is used for classification.

The reason for making this experiment is that all the captured face can be directly obtained by traditional face detection method and then face recognition can be per-

Table 4.11: Accuracy (%) on FERET [1] database from our cropped images.

Method \ Probe	Fb	Dup I	Dup II
LBP	83.5	44.5	28.4
1DLPMS-B	86.2	47.2	30.5
G1DLPMS-B	85.2	45.9	28.8
LTP [26]	87.6	48.3	31.3
LDP [9]	88.3	47.4	30.5
1DLPMS-T	89.3	49.9	33.9
G1DLPMS-T	91.4	53.7	39.9
1DLPMS-Q	89.5	50.3	35.6
G1DLPMS-Q	93.4	53.3	39.0
weighted LCBPMS	97.3	69.3	62.6
weighted LLCP	98.8	75.3	73.2

formed without any normalization according to the known coordinates of eyes and mouth. And in order to compare to some start-of-the-art methods, same experimental setup in [9] is followed. Also, Gabor feature is considered and the results is show in Table 4.12. Prefix G_ means Gabor feature is used in that corresponding method. In this evaluation, 1,199 persons with a total of 14,051 gray-level faces are used in the FERET dataset. There are many variations for each person, such as different illumination condition, various facial expressions, varied pose angles and so on. In this study, only the frontal faces are considered and all the faces can be categorized into the following five sets: 1) Fa set, which is considered as a gallery set. There are totally 1,196 frontal faces with each one per person. 2) Fb set, which concludes 1,195 faces. Compared to Fa set, the facial expression is various for each people. 3) Fc set, which concludes 194 faces. The key variation between this set and the previous two ones is that the faces in Fc set are captured in various illumination scenarios. 4) Dup I set, which concludes 722 faces. Different from the above three sets, all the faces in this set are captured some times later. 5) Dup II set, which concludes 234 faces, is a subset of Dup I set containing the faces which are captured at least one year later compared to Fa set. And in our study all facial images are normalized to

80 × 88 pixels and the procedure of the FERET evaluation principle [1] is followed.

The total performance on FERET database is listed in Table 4.12 (Note that: in LCBP, LCBPMS and LLCP, just histogram intersection is used for classification and our proposed WHSEMD classifier is applied in Weighted LCBP, Weighted LCBPMS and Weighted LLCP methods). From Table 4.12, we can see that Curvelet feature is effective in our proposed methods, especially in Dup I and Dup II set since Curvelet has good characteristics in space frequency, space position and direction selectivity. LCBP and Weighted LCBP mean that when encoding the neighborhood of the patterns, circle is used. We can also see that multi-scans neighborhood based encoding is effective in the Curvelet based frequency domain.

Fig.4.10 investigates how the performance will be changed when R and resolution of images are varied (1DLPMS-B and fb probe set is used). Three kinds of resolution and four kinds of R are selected. From this figure, we can see that higher performance can be obtained by better resolution. For small resolution images, a little larger R can get rapid increasing in accuracy (R from 1 to 3 in 50 × 50 resolution) while performance is almost same for different R in large resolution images. (R from 2 to 4 in 100 × 100 resolution). The reason may be that in small resolution, larger R can capture more local information, while in large resolution, with the increasing of R, little local information will be added.

In order to determine how the parameters affect the feature descriptors proposed in chapter 3 during the final recognition rate, several experiments are evaluated on Fb and Dup I databases. Fig.4.11 and Fig.4.12 show the recognition rates change with the size of codebook K and different sampling methods (Here, the region size is fixed to 8 × 8 and just histogram intersection is used for classification). From these two figures, we can find that the performance becomes better with the increase of K. When K is from 256 to 512, the recognition rate stay stable and just 0.5% improved from 128 to 256. So for the sake of trade-off between precision and computation cost,

Table 4.12: Accuracy (%) on FERET [1] database by standard principle.

Method \ Probe	Fb	Fc	Dup I	Dup II
PCA	85	65	44	22
Fisherface	94	73	55	31
HOG-EBGM [59]	96	82	60	52
SIFT Grid [60]	94	35	53	36
SURF	91	58	64	63
AAM [61]	96	93	73	77
DRDWT [62]	98	93	76	79
LBP [6]	91	65	53	38
ALBP [7]	93	67	59	41
DLBP [28]	89	69	61	43
LGBP [10]	94	97	68	53
Weighted LGBP [10]	98	97	74	71
HGPP [11]	97	98	77	76
Weighted HGPP [11]	97	99	79	78
LDP [9]	92	88	63	60
G_LDP [9]	98	99	80	80
LLGP [12]	97	97	75	71
Weighted LLGP [12]	99	99	80	78
POEM-HS [8]	98	96	78	76
Retina filter + POEM-HS [8]	98	99	80	79
IDLPMS-B	94	89	69	67
G_1DLPMS-B	97	97	78	75
IDLPMS-T	96	93	73	71
G_1DLPMS-T	99	99	83	82
IDLPMS-Q	98	97	75	76
G_1DLPMS-Q	99	99	82	84
LCBP	97	97	73	71
Weighted LCBP	99	98	79	75
LCBPMS	98	97	76	74
Weighted LCBPMS	99	98	82	78
LLCP	98	98	80	75
Weighted LLCP	99	99	85	82

$K = 128$ is used in the following. For the sampling methods, sample 2 is better than sample 1 while the combination of them can achieve the best result.

Next experiment is based on the change of region size and sampling methods while the K is fixed to 128. The results are shown in Table 4.13 and Table 4.14.

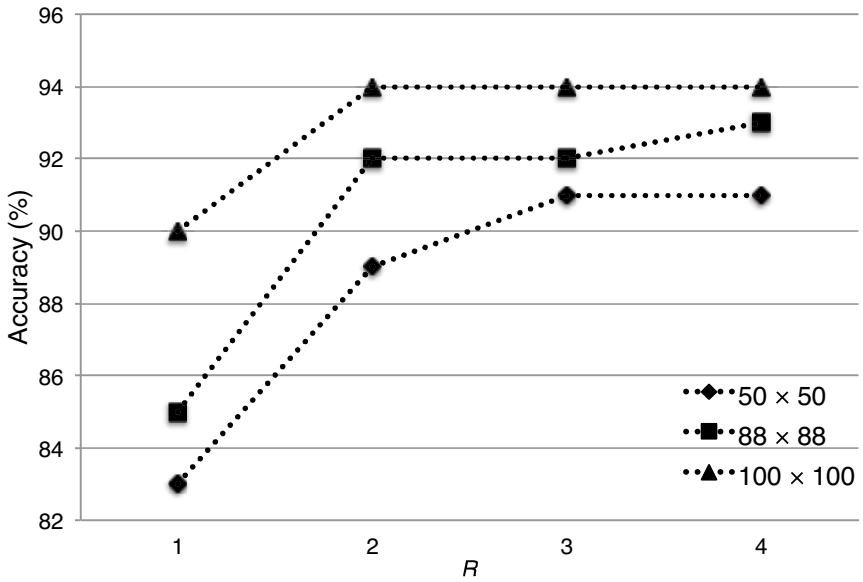


Fig. 4.10: Performance when R and resolution of images are varied.

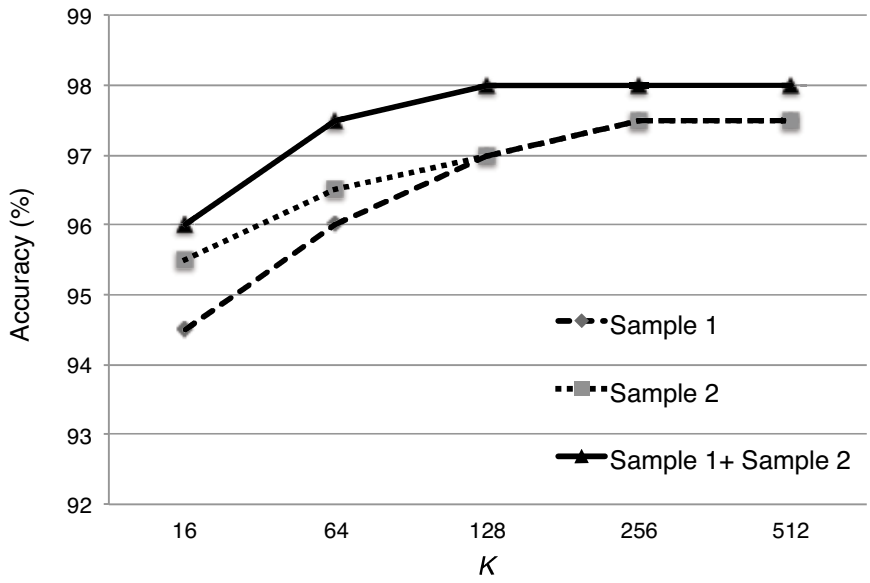


Fig. 4.11: Performance of LLCP on Fb with different size of codebook and sampling methods.

From these two tables, we can seen that small or large region size can decrease the performance, especially with larger one. When the region size is about 8×8, the

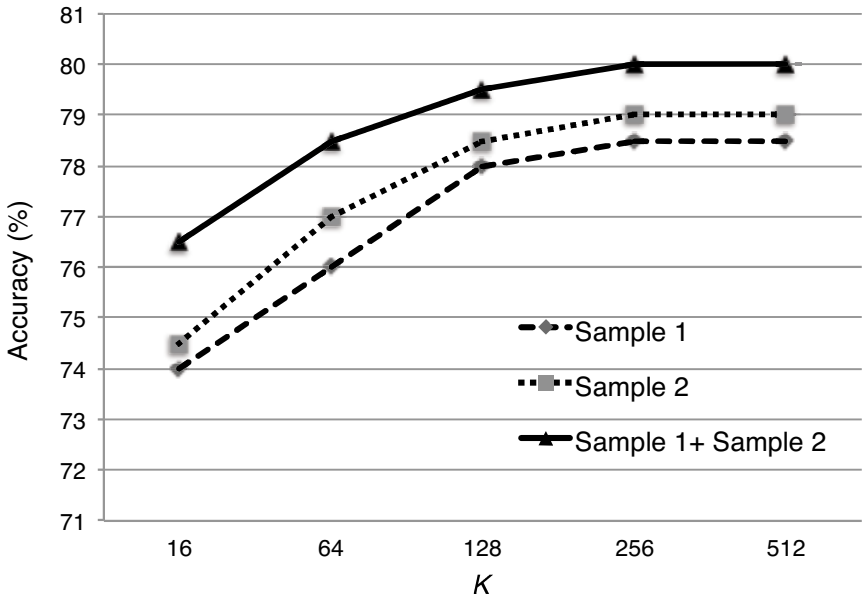


Fig. 4.12: Performance of LLCP on Dup I with different size of codebook and sampling methods.

best accuracy can be obtained. The possible reason may be that larger region size can not preserve enough spatial information in the facial images while the patterns in smaller region can not discriminate effectively. For the sampling methods, same as the previous one, the combination can get the toppest rate.

Table 4.13: Performance of LLCP on Fb with different region size and sampling methods.

Region Size \ Sample Method	Sample 1	Sample 2	Sample 1 + Sample 2
4 × 4	99.5	96.0	97.0
4 × 8	97.0	97.5	98.0
8 × 8	97.5	97.5	98.0
8 × 11	97.0	97.5	98.0
10 × 11	96.5	97.0	97.5
20 × 22	95.5	96.0	96.5

The following experiment is evaluated on the size of training set and different sampling methods and the results are shown in Fig.4.13 and Fig.4.14. We can see

Table 4.14: Performance of LLCP on Dup I with different region size and sampling methods.

Region Size \ Sample Method	Sample 1	Sample 2	Sample 1 + Sample 2
4×4	77.5	77.0	79.0
4×8	78.0	78.5	79.0
8×8	78.0	78.5	79.5
8×11	76.0	77.5	79.0
10×11	73.0	75.0	76.0
20×22	70.0	71.0	71.5

that only small database can be conducted robust recognition rate while larger ones do not increase the final performance.

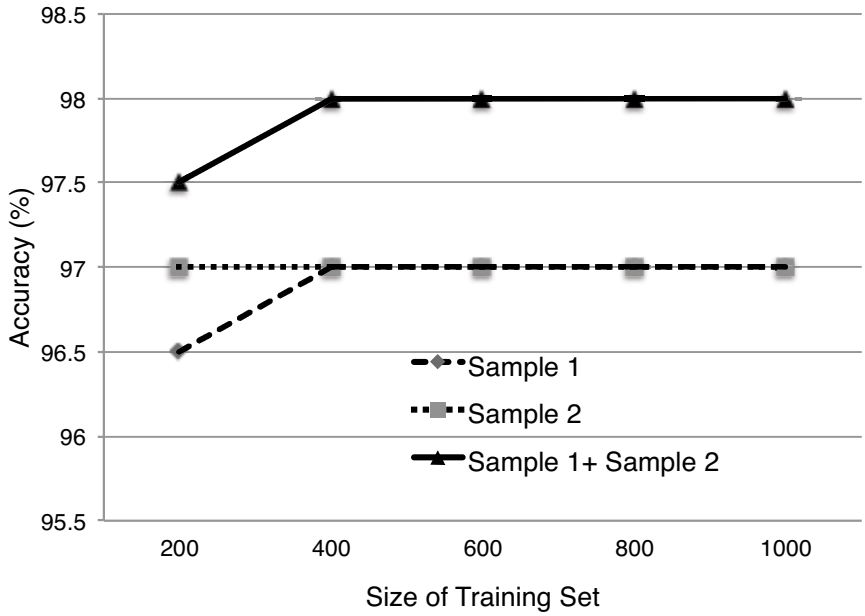


Fig. 4.13: Performance of LLCP on Fb with different size of training set and sampling methods.

The next experiment is designed to judge whether our multi-mapping strategy is useful or not in Curvelet space. The evaluation based on Dup I is list in Fig.4.15. We can see that a little larger β can improve our performance which is same as our thinking. And if β is so large, that means the input patch is also encoded by some

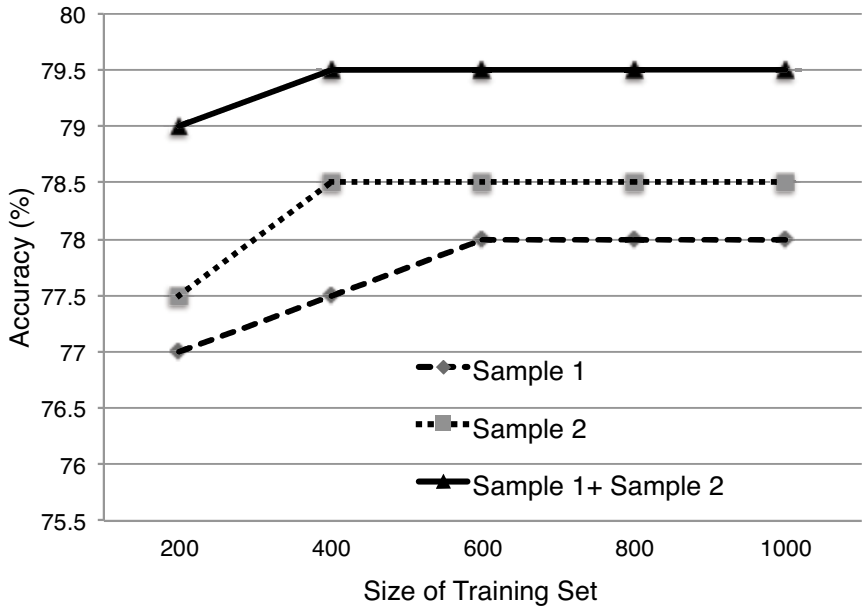


Fig. 4.14: Performance of LLCP on Dup I with different size of training set and sampling methods.

dissimilar patterns which can confuse the distribution of final histogram and decrease our final result.

Another experiment is designed to judge how the recognition rate changes with variation of image size. The evaluation result based on FERET database is list in Fig.4.16 where the combined sampling method is used. We can see that when the image size is larger than 80×88, the performance is almost same, while just a little improved in Dup I and Dup II subsets with the image size from 100×100 to 128×128. Thus, image size with 80×88 is sufficient for this kind evaluation on this database.

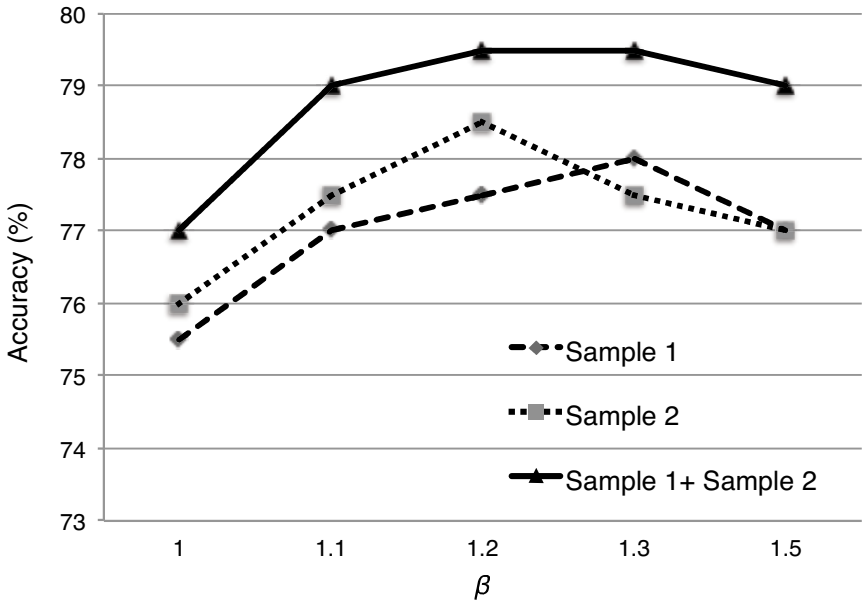


Fig. 4.15: Performance of LLCP on Dup I with different parameter β in Eq. (3.6).

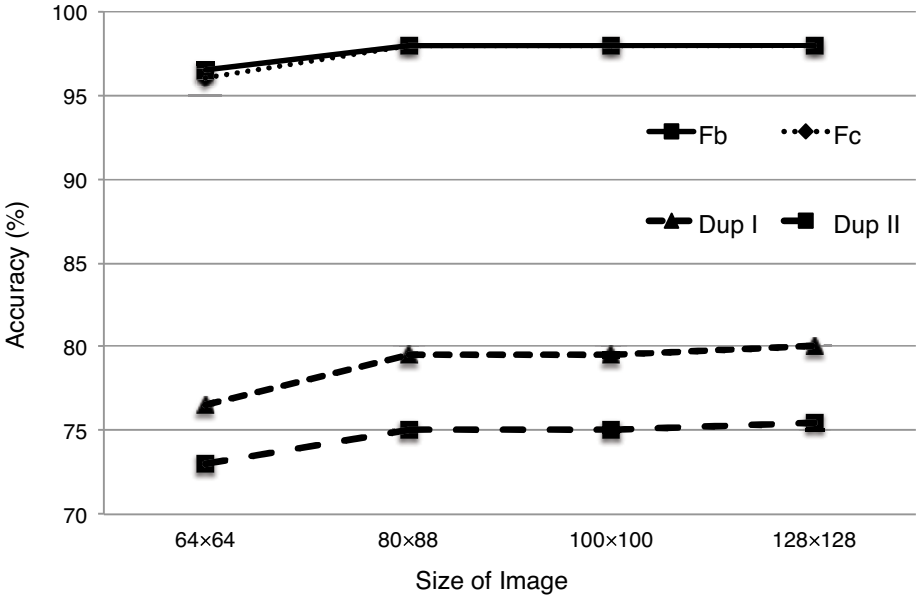


Fig. 4.16: Performance of LLCP on FERET database with different image size.

In the following evaluation, we chose a subset of the FERET database to evaluate our proposed LDA feature projection methods. It includes 900 samples of 150 persons

(each person has six samples). For each individual, there are two or three frontal faces with different lighting conditions and facial expressions, and the remaining faces are pose variations with angle from -15° to +15°. Each face is cropped into resolution of 100 by 80 and no preprocessing method is applied. Fig.4.17 illustrates six cropped images of one person. During this experiment, there are totally 450 faces (3 faces per person) are randomly used in the training stage. At the same time, these training faces are also treated as a gallery set while the left 450 faces are treated as a probe set. The average precision is recorded by computing the average performance of recognition rates across 5 runs.

Fig. 4.17: Some samples from FERET database.

Several experiments are conducted to judge whether our proposed methods are powerful or not when some other feature spaces are applied. Here, we selected three local patterns: uniform LBP [6] (the number of sampling points is set 8) and our proposed 1DLPMS-B [63] in chapter 2 (Here, eight kinds of scans are used and the number of sampling points for each scan is set 6) and LCBPMS [64] in chapter 3. LBP first locates a block to each pixel in the image through thresholding a 3 ×3 neighborhood points with the center one to encode the final pattern with binary number. Second, all the binary numbers generate a histogram to be treated as the feature vector of this face. In 1DLPMS-B, multi-scans are used to capture the different spatial information on the facial image. Compared to LBP, which only uses a circle to encode the neighborhood pixels, multi-scans can keep more spatial information, reduce the effect of illumination and noise problem and 1DLPMS-B is

scale and rotation invariant. LCBPMS is the method which extend 1DLPMS-B into Curvelet space to reduce the effect of illumination, noise, facial expression, pose and so on. First, each facial image is equally divided into 80 regions, and then the above mentioned local patterns are applied into each region. The average recognition rate by LBP is 85.6% while 88.5% and 92.9% for 1DLPMS-B and LCBPMS, respectively. Fig.4.18 shows some classification results corresponding to LBP combined with some feature reduction methods and the results by combing 1DLPMS-B or LCBPMS and related feature reduction methods are shown in Fig.4.19 and Fig.4.20, respectively.

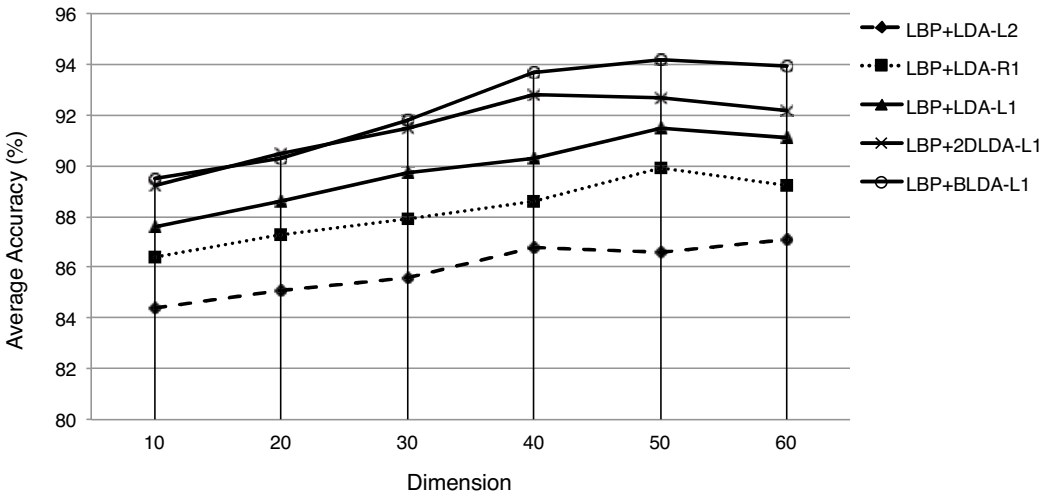


Fig. 4.18: Recognition rate by combine LBP and some feature reduction methods from FERET database.

From these three figures, we can see that the feature projection methods are very effective and efficient in the feature spaces, which are extracted by local patterns. More compact and discriminative feature can be obtained by the feature selection. Our proposed methods can improve accuracy by about 8 percent while 4 percent for LDA-R1 and 1 percent for LDA-L2, respectively. Note that the feature vector was firstly converted into two dimensions when 2DLDA-L1 and BLDA-L1 was applied.

In the next estimation, the average reconstruction errors for FERET database

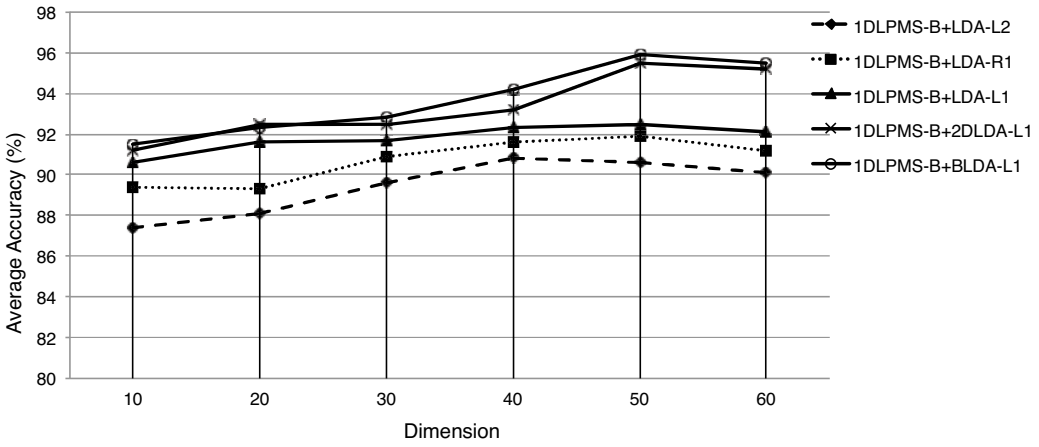


Fig. 4.19: Recognition rate by combine 1DLPMS-B and some feature reduction methods from FERET database.

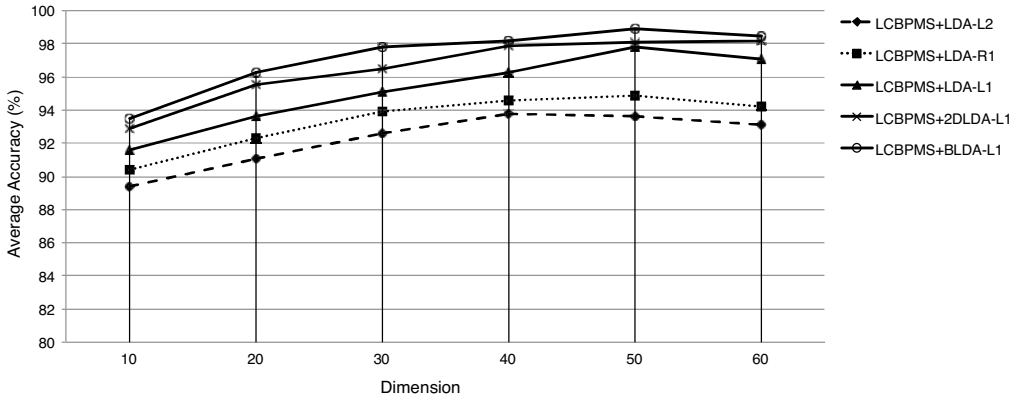


Fig. 4.20: Recognition rate by combine LCBPMS and some feature reduction methods from FERET database.

with LBP, 1DLPMS-B and LCBPMS are shown in Fig.4.21, Fig.4.22 and Fig.4.23 with various numbers of extracted features. From these figures, we can also see that our proposed methods are more suitable for reconstruction in local patterns based feature space, while BLDA-L1 performs best compared to other approaches.

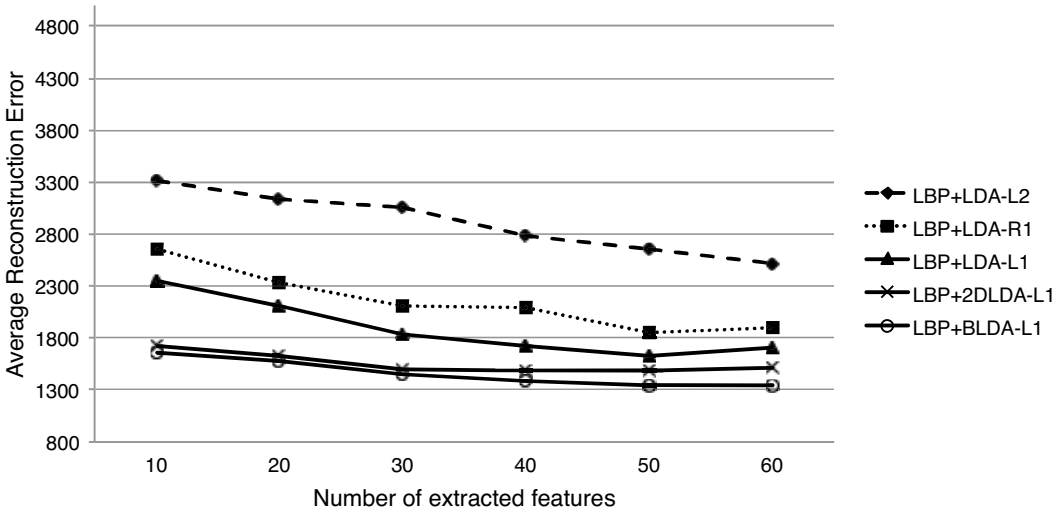


Fig. 4.21: Average reconstruction errors for FERET database by LBP.

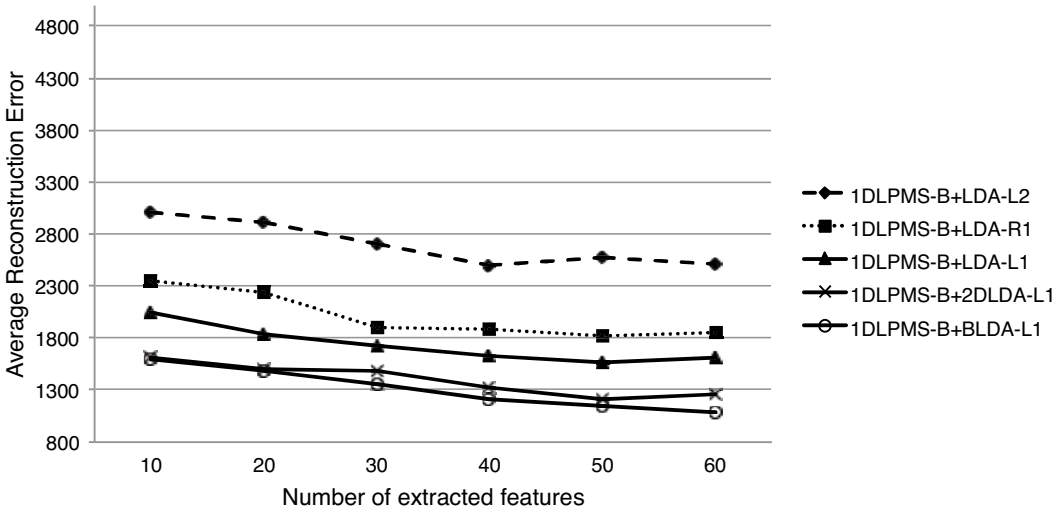


Fig. 4.22: Average reconstruction errors for FERET database by 1DLPMS-B.

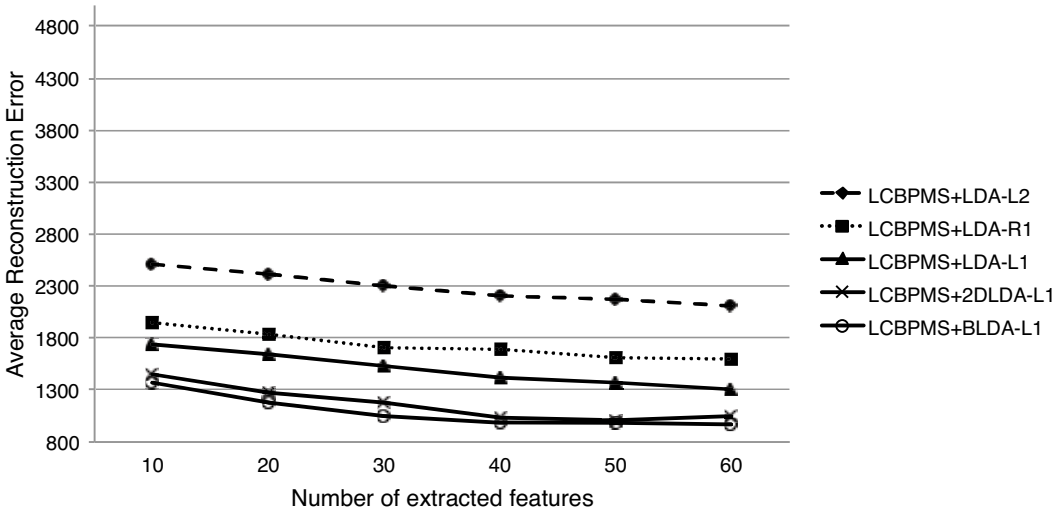


Fig. 4.23: Average reconstruction errors for FERET database by LCBPMS.

4.5.3 FRGC Database

The FRGC [56] database consists of over 50,000 frontal facial recordings of more than 4,00 subjects with frontal views at various facial expressions and illumination conditions. For the experiments reported in this section, 200 different individuals are randomly selected from this database and each subject has 8 images. Then there are totally 1600 images in our experiments. All the images are manually cropped and normalized into 80×88 pixels and divided into 8 by 8 regions in our study. Some examples are shown in Fig.4.30(b).

In this evaluation, some $D_{num}(D_{num} = 1, 2, 3, 4)$ image(s) of each person will be randomly selected for training, while the left images for testing. To compare our method with LBP, LGBP, five tests are evaluated with a different number of training sets and mean rate is recorded. Table 4.15 shows the accuracy. From this table, we can see again that Curvelet based local patterns are more effective than the traditional local patterns and Weighted LLCP is the outstanding one.

Table 4.15: Accuracy (%) on FRGC- D_{num} (1,2,3,4) for gallery and others for probe.

Method \ D_{num}	1	2	3	4
LBP [6]	75	82	88	94
LGBP [10]	79	05	91	95
LCBP	82	89	92	97
Weighted LCBP	85	91	94	98
LCBPMS	84	91	93	97
Weighted LCBPMS	86	92	95	98
LLCP	89	93	95	99
Weighted LLCP	91	95	97	99

4.5.4 AR Database

There are totally 126 persons with over 3200 frontal color faces in the AR dataset [2], and each person consists of 26 different images with the variation of various occlusions, illumination scenarios and facial expressions. For every person, the faces

are captured in two varied sessions separated by two weeks with 13 faces in each session. For the experiments reported in this section, 60 different individuals are randomly selected from this database. Then there are 1560 images in our setup. All the samples are manually cropped and normalized into 80 by 60. Some samples of one person are illustrated in Fig.4.24.

Fig. 4.24: Some samples from AR database.

In this evaluation, the recognition performances of the different approaches on AR database are compared. Six samples of every individual are randomly chosen as gallery (training images), and the left ones are used for probe (testing images). In our study, 5 times are performed to randomly select the training set and the average recognition rates are recorded. Some classification results are listed in Fig.4.25, where we can see that our proposed method has higher performance than LDA-L2 and LDA-R1. In general, LDA-L1 can obtain about 10 percent or 20 percent than LDA-R1 and LDA-L2, respectively. And the average number of iterations for LDA-L1 is 7.5 while 25.3 for LDA-R1. Thus, we can see clearly that LDA-R1 takes much more computation cost to achieve convergence in larger dimensional input space, such as face recognition application, than LDA-L1. Base on this evaluation, our

proposed methods are more effective and efficient than the traditional approaches to solve facial expression, illumination or occlusions issues.

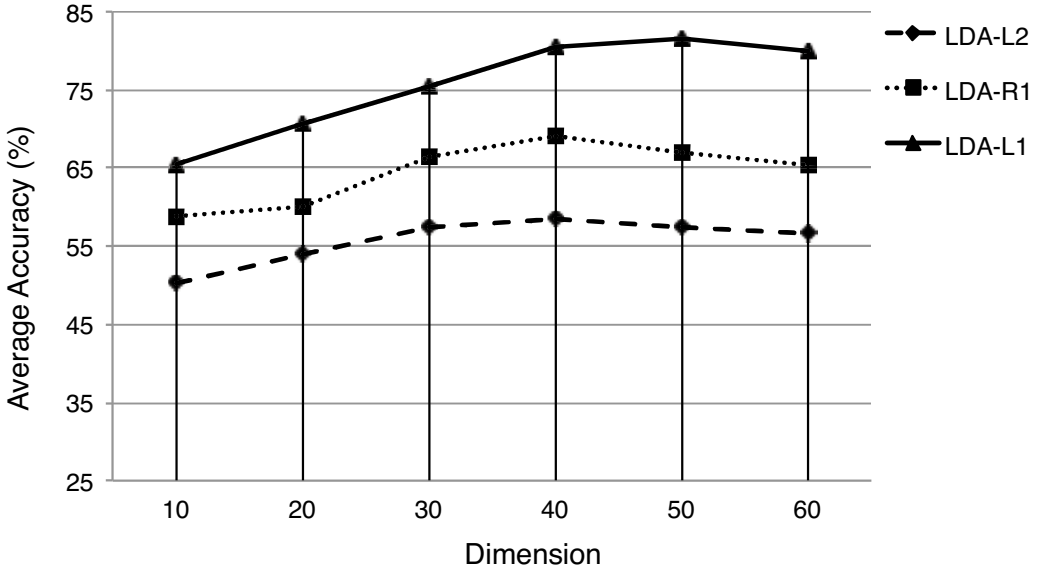


Fig. 4.25: Classification results on AR database.

In Fig.4.26, only low-dimensional space is focused on since we want to make a comparison of the most discriminant features for our proposed method and some related algorithms. Same as Fig.4.25, our proposed method can extract more discriminant features.

Finally, the average accuracy and time cost for training on AR database is concluded in Table 4.16, and our proposed methods are superior to the traditional approaches.

Table 4.16: Accuracy and computation cost on AR database.

Method	Accuracy (%)	Average number of iterations	Average time (s)
LDA-L2 [16]	58.6	/	/
LDA-R1 [23]	69.1	25.3	87.8
PCA-L2 [14]	55.2	/	/
PCA-L1 [21]	73.2	/	/
LDA-L1	81.6	7.5	63.9

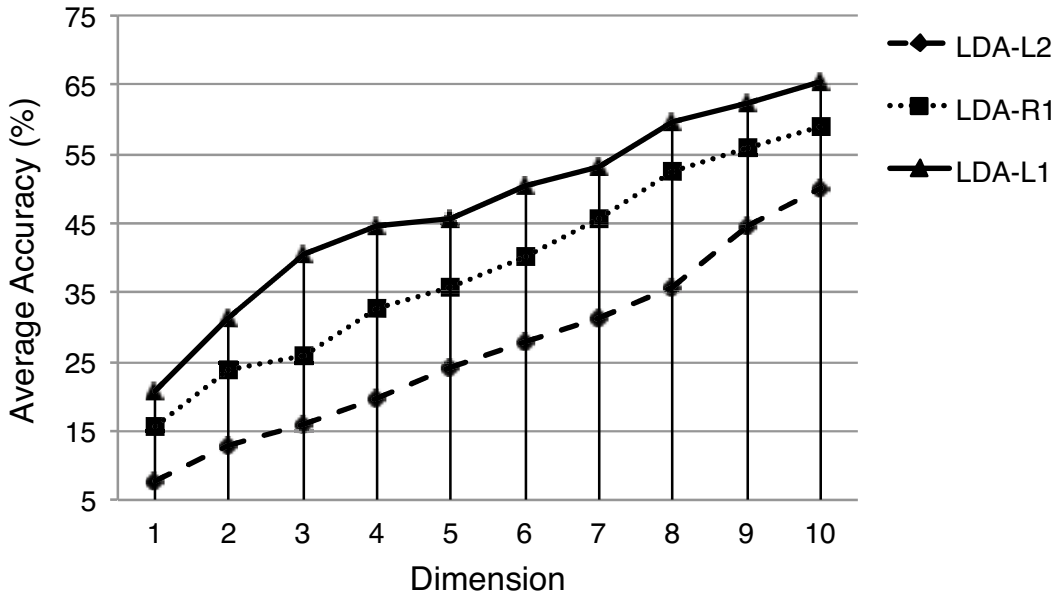


Fig. 4.26: Classification results for small dimension on AR database.

In the next evaluation, more accurate results of different approaches related to L1-norm based algorithms are compared on AR dataset . In the learning stage, six samples per person are randomly selected, while the left ones are applied in the testing step. The algorithms are run 10 times with randomly selecting the learning set and the average recognition rates are recorded in final. Some more classification results are listed in Fig.4.27, where we can see that our proposed methods have higher performance than LDA-L2 and LDA-R1, especially for 2DLDA-L1 and BLDA-L1. (Note that in 2DLDA-L1 and BLDA-L1, the Dim. is equal to $t1 \times t2$)

In Fig.4.28, we focus on only low-dimensional spaces because we want to make a comparison of the most discriminant features for our proposed methods and some related algorithms. Same as Fig.4.27, our proposed methods can extract more discriminant features.

The average reconstruction errors for AR database are shown in Fig.4.29 with various numbers of extracted features. From this figure, we can see that the average

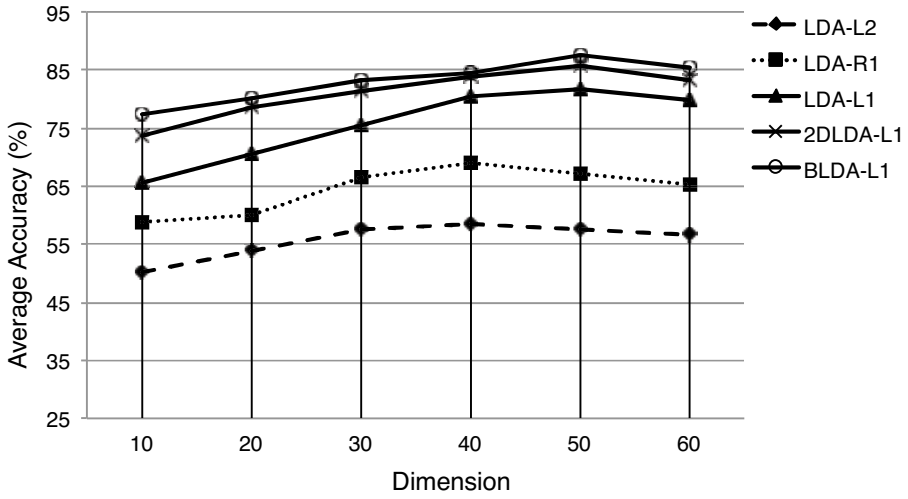


Fig. 4.27: More classification results on AR database.

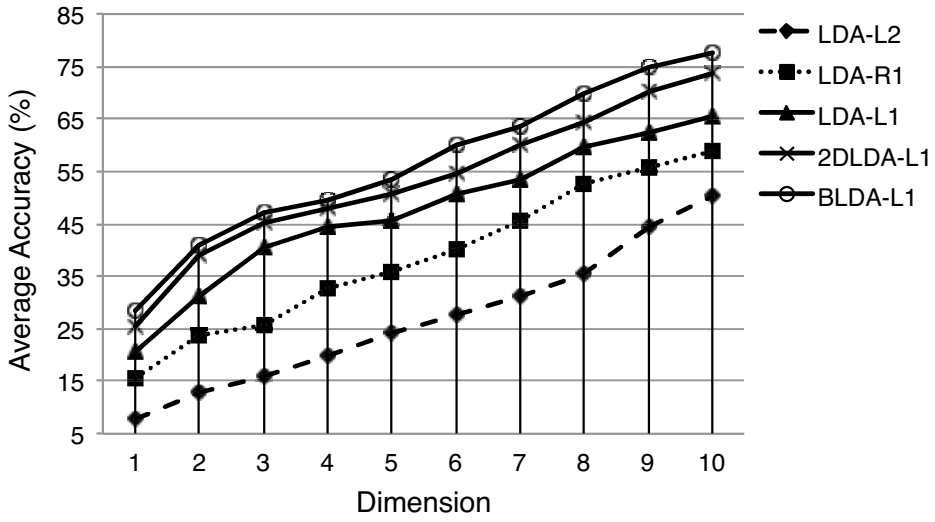


Fig. 4.28: More classification results for small dimension on AR database.

reconstruction errors of our proposed methods are much smaller than the traditional approaches while BLDA-L1 is a little better or comparable than 2DLDA-L1.

Finally, the total accuracy on AR database is concluded in Table 4.17, and our proposed method can improve about 6% compared to the the state-of-the-art approach.

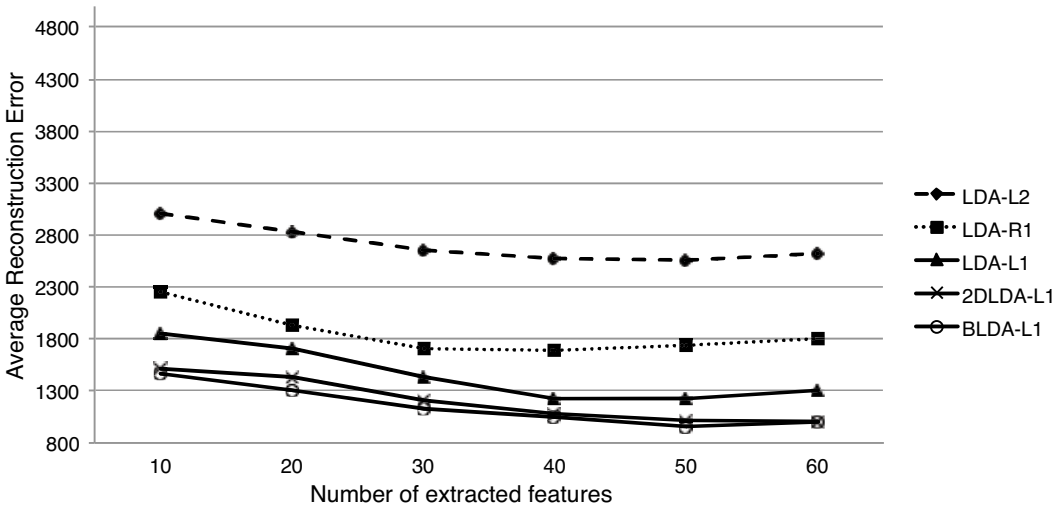


Fig. 4.29: Average reconstruction errors for AR database.

Table 4.17: Accuracy on AR [2] database.

Method	Accuracy (%)
LDA-L2 [16]	58.6
LDA-R1 [23]	69.1
2DPCA-L1 [22]	82.2
LGBP+LDA-R1	88.7
LGBP+2DPCA-L1	89.3
LDA-L1	81.6
2DLDA-L1	85.7
BLDA-L1	87.5
1DLPMS-B +LDA-L1	88.3
1DLPMS-B +2DLDA-L1	92.1
LCBPMS +LDA-L1	91.6
LCBPMS +2DLDA-L1	95.9

4.6 Gender Estimation

In this experiment, two large databases-Labeled Faces in the Wild (LFW) [55] and Face Recognition Grand Challenge(FRGC) [56] are used. Some samples are shown in Fig.4.30.

LFW dataset is designed for analysis the faces under unconstrained conditions. This dataset is consist of more than 13,000 pictures of face samples that are all

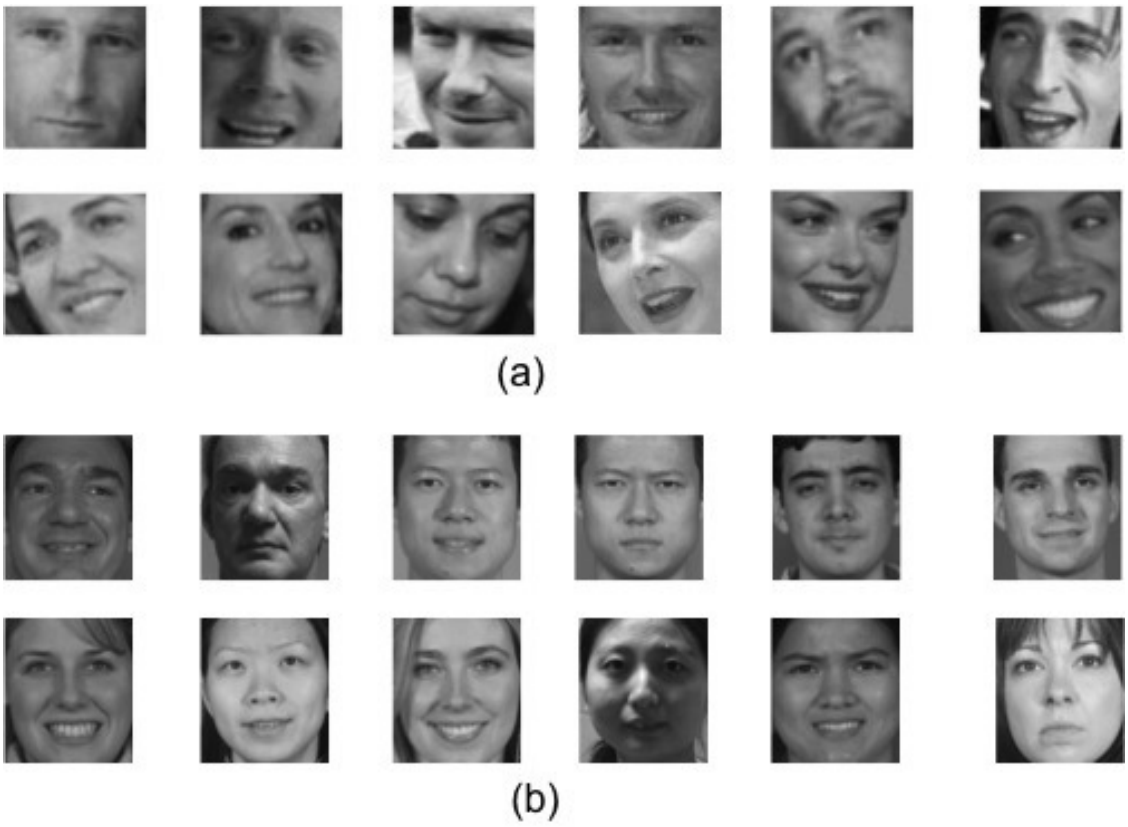


Fig. 4.30: Some samples for gender estimation. (a) LFW; (b) FRGC.

obtained from the website. Every face sample is labeled by the name of the people captured, and there are 1,680 persons who have two or three samples on this dataset. In our study, all the face samples are detected by the well-known facial detector proposed by Viola and Jones. We randomly select 2000 males and 1000 females for training and use another 2000 males and 1000 females for testing. For FRGC database, 1000 males and 1000 females are selected for training and other 1000 males and 1000 females are used for probe. These images are cropped to 64×64. The most difference between the two databases is that the images on FRGC are all frontal faces while not on LFW database. A support vector machine (SVM) classifier is selected as the classifier in our gender estimation system since it is well known and powerful

in the area of statistical learning theory and has been successfully used in gender estimation. In literature, SVM seems effective and efficient to deal with a binary classification issue with the concept of structural risk minimization (SRM) [65], [66] to find the linear decision hyperplane optimally. In our system, three kernel f

le 4.18. $\dot{g}$ is equal 3 in our system. $x^{'}$ and $y^{'}$ are feature vectors and $x^{'}, y^{'} \in \mathbb{R}^{m}$.

Table 4.18: Kernel functions using in the system.

Dot product	$k(x^{'}, y^{'}) = x^{'} \cdot y^{'}$
Polynomial	$k(x^{'}, y^{'}) = (x^{'} \cdot y^{'} + 1)^{\dot{g}}$
RBF	$k(x^{'}, y^{'}) = exp(-\frac{\lvert x^{'} - y^{'} \rvert^2}{2\delta^2})$

The first experiment is to evaluate how the precision changes with variation of the number of regions. Fig.4.31 and Fig.4.32 show the results based on LFW and FRGC databases, respectively. $0, 1, 2$ means the kernel used in SVM is dot product, polynomial and RBF, respectively. From these two figures, we can see that when the number of regions is set about 64, the performance achieves best, and decreases when the number of regions is larger or smaller. The reason may be same as discussed in face identification problem.

Table 4.19: Accuracy (%) on LFW database.

Method \ Kernel type	0	1	2($\delta = 30$)
LBP	85.5	85.0	86.0
LGBP	87.0	87.5	88.5
1DLPMS-B	86.5	85.6	86.4
1DLPMS-T	86.6	90.0	86.7
1DLPMS-Q	89.0	91.6	89.6
LCBP	91.5	89.5	92.0
LCBPMS	92.0	91.5	92.5
LLCP	93.5	95.0	93.5

The total experimental results are shown in Table 4.19 and Table 4.20, respectively. Here, the number of regions is set to 64. From these two tables, we can

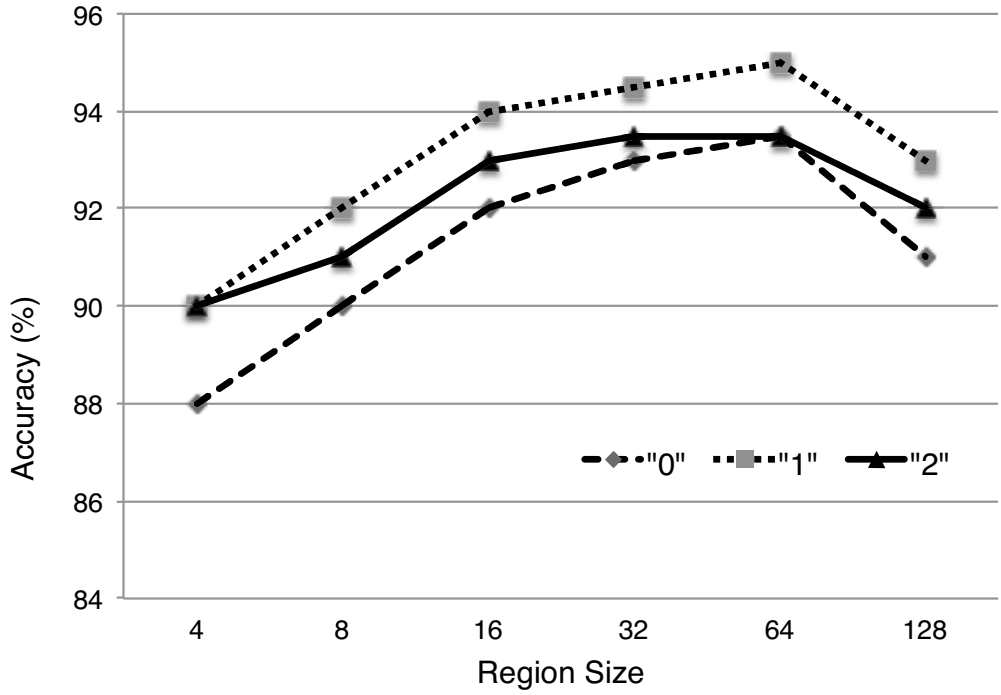


Fig. 4.31: Accuracy of LLCP on LFW database with different number of regions.

T .20: Accuracy (%) on FRGC database.

Method \ Kernel type	0	1	2($\delta = 30$)
LBP	93.5	95.0	92.5
LGBP	94.5	96.0	94.0
1DLPMS-B	93.9	95.2	94.5
1DLPMS-T	96.9	96.9	93.7
1DLPMS-Q	97.1	96.3	95.2
LCBP	97.0	97.5	94.0
LCBPMS	97.5	98.0	96.5
LLCP	98.0	98.5	97.5

see that our proposed methods can get more discriminant features about male and female than LBP and LGBP under different variations. And frontal faces are more useful for gender estimation than profile.

Next evaluation is tested on male and female set, respectively. We try to find which gender is harder to be recognized compared to the other one. The results

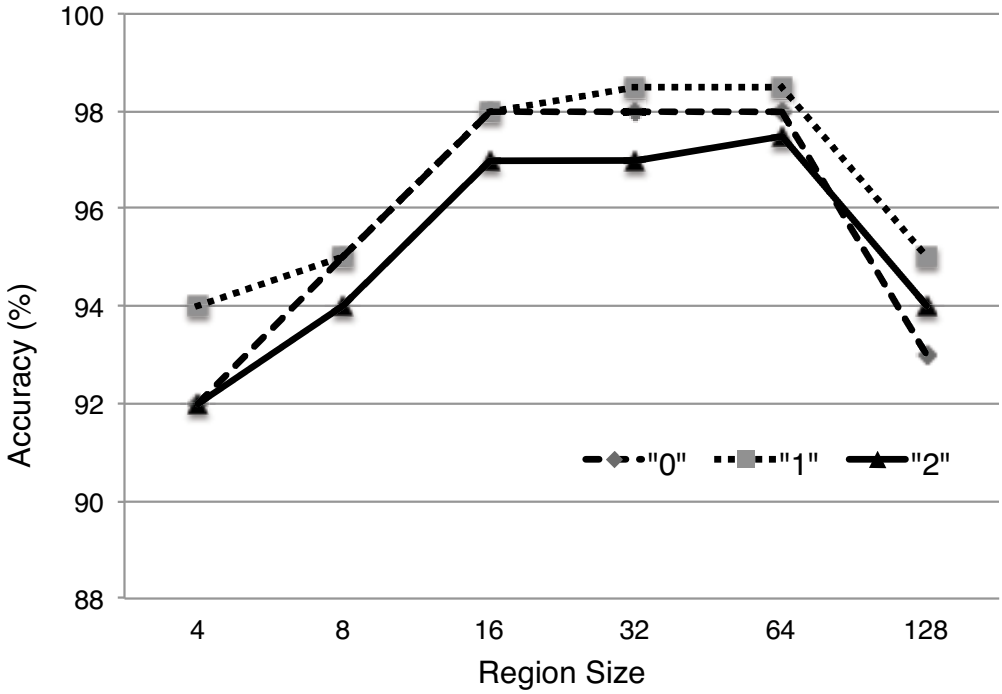


Fig. 4.32: Accuracy of LLCP on FRGC database with different number of regions.

evaluated on LFW and FRGC are list in Table 4.21 and Table 4.22, respectively. From these two tables, we can see that the recognition of male is much higher than female on LFW database, while the recognition rate of female is almost same as male on FRGC database. So male is expected easier to be recognized in our real life while comparable with female if just frontal face is used to estimation.

Table 4.21: Accuracy (%) about male and female on LFW database, respectively.

Method \ Kernel type	0		1		2	
	Male	Female	Male	Female	Male	Female
LBP	94.7	64.0	92.1	76.9	95.0	65.2
1DLPMS-B	94.3	71.1	91.9	81.6	95.5	68.3
1DLPMS-T	94.6	70.7	92.9	84.2	95.6	69.0
1DLPMS-Q	91.0	85.0	94.2	86.6	92.6	76.7
LCBPMS	95.2	85.1	94.6	88.6	97.6	94.0
LLCP	97.3	89.0	96.2	89.6	98.6	93.7

Table 4.22: Accuracy (%) about male and female on FRGC database, respectively.

Method \ Kernel type	0		1		2	
	Male	Female	Male	Female	Male	Female
LBP	92.0	94.3	93.5	96.6	90.6	94.2
1DLPMS-B	91.6	96.1	92.6	97.8	93.2	95.7
1DLPMS-T	96.4	97.3	96.1	97.7	92.5	93.8
1DLPMS-Q	97.0	97.1	96.3	96.3	92.6	95.5
LCBPMS	98.2	97.9	98.2	97.3	97.9	94.2
LLCP	98.3	98.0	98.9	98.1	98.2	95.3

4.7 Facial Expression Recognition

From the previous evaluations, we can see clearly that the proposed 1D local features are powerful to deal with the face identification and gender estimation problems. Specially, the performance of our approaches is better than the stat-of-the-art with the advantage of low time cost. However, in the above approaches, the regions are equally divided in the facial images. The disadvantage is that the features are highly related to the regions since these features are extracted from the fixed size of region and location. Thus, in order to get more reasonable features, multi-scale and overlapping schemes are applied in the following study. The different sizes of windows are shifted over the facial images to obtain more regions. As a result, more detail and powerful description of facial features can be got [67]. For the sake of extracting more discriminative features from the lots of histograms introduced by multi-scale and overlapping scheme, boosting learning [68] is applied to obtain more significant histograms. In Ref. [69], Zhang et al. applied boosting LBP-based classifiers to analyze the faces, where the discriminative feature is treated as the distance between the corresponding LBP facial histograms, and in final AdaBoost is applied to get a set of most powerful features. In our application, same scheme and JAFFE database [57] are studied.

The JAFFE dataset as illustrated in Fig.4.33 is consist of 213 pictures of 7 facial expressions with one neutral and six basic facial expression captured from ten Japanese females [57]. Each image is grouped into six emotions by rating from 60 Japanese persons. And for one subject, there are two to four samples for every expression. 15 images for each expression are selected as gallery randomly while others are for probe. SVM is applied for classification and the resolution of all images is 64 by 64.

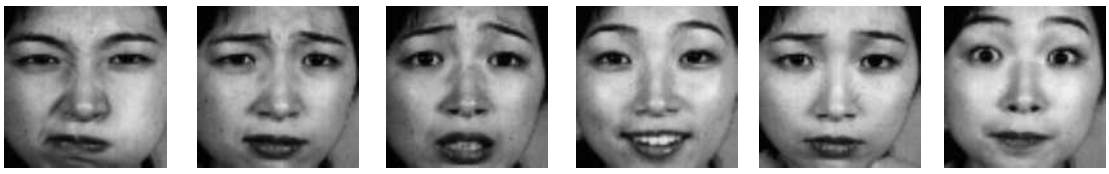


Fig. 4.33: Samples from JAFFE database. The facial expression from left to right is angry(AN), disgust(DI), fear(FE), happy(HA), sad(SA) and surprised(SU).

In the feature-extracting step, AdaBoost can extract the most discriminative and powerful parts of facial expression among all the computed histograms. For the classifier, just a weak one called histogram-based template matching as same one as in Ref. [69] is used.

In this multi-scale and overlapping scheme, there are totally 625 histograms/regions can be obtained for each facial image. Specially, the resolution of the scaled windows is from 16 $\times$ 16 pixels to 20 $\times$ 20 with the scaling step of 4 pixels, while the window is also shifted at the facial sample by the step of 4 pixels. The effective and powerful histograms can be obtained by AdaBoost algorithm. We plot in Fig.4.34 and Fig.4.35 the spatial localization of the 6 sub areas that are related to the top 6 histograms chosen by Boosted-LBP and Boosted-1DLPMS-B for two comparable expressions, respectively. Here, just the first four kinds of scans are used, since the remaining ones are the reverse versions of the first four kinds of scans.

Fig. 4.34: The sub-regions (LBP histograms) selected by AdaBoost for each emotion.

From the two figures, we can see that Boosted-1DLPMS-B can select more powerful regions than Boosted-LBP. Take AN vs DI for an example, in Fig.4.36, one region is selected between right eye and nose, and another region is selected in the bottom right cheek. Both of these two regions are not so varied in AN and DI facial expression. But, in Fig.4.35, all six regions are located in eyebrows, eyes and mouth where the difference between AN and DI facial expression is significant.

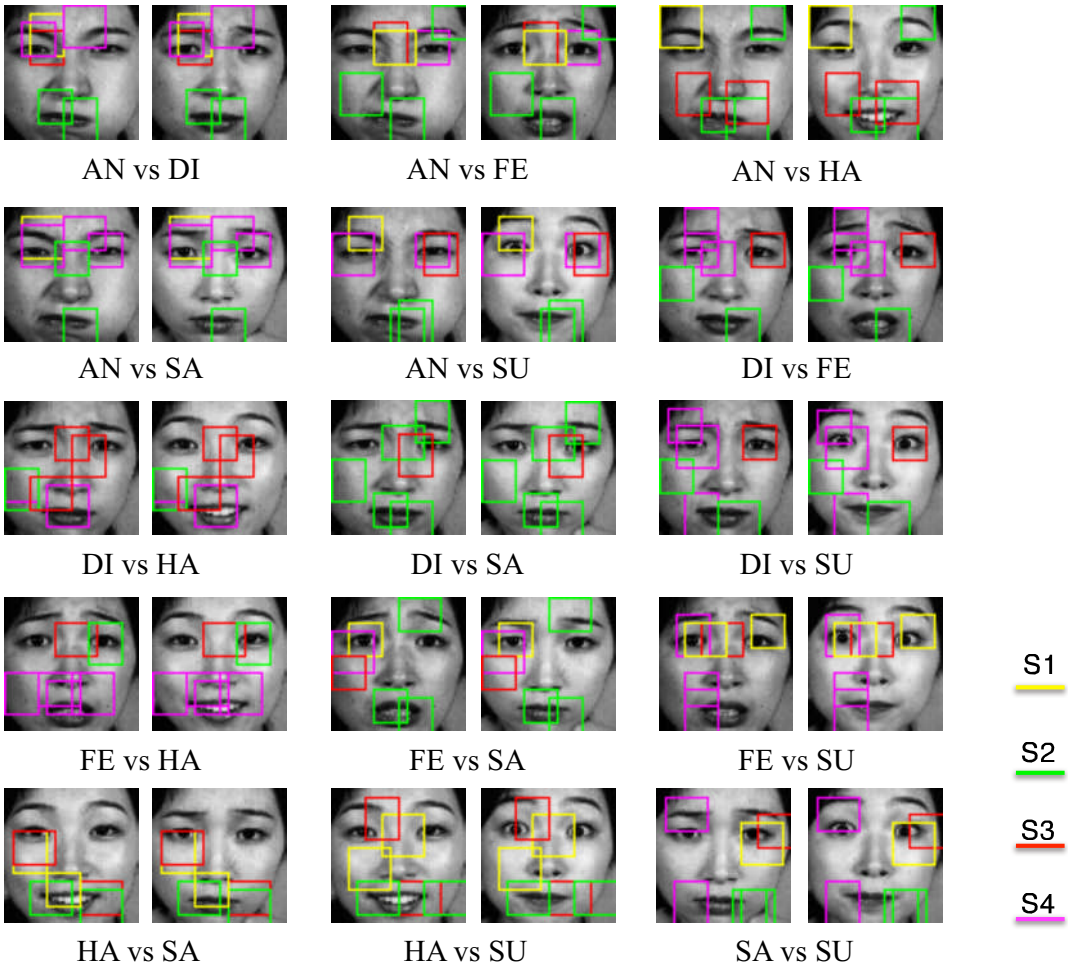


Fig. 4.35: The sub-regions (1DLPMS-B histograms) selected by AdaBoost for each emotion.

Table 4.23: Average accuracy about facial expression.

Method	Accuracy (%)
LBP	82.4
Boosted-LBP	86.7
1DLPMS-B	87.2
Boosted-1DLPMS-B	91.3

The comparison of two facial expression accuracy is illustrated in Fig.4.37 and average precision is show in Table 4.23. In Fig.4.37, one against one strategy by SVM classifier is used. For example, in AN vs DI condition, all the probe images

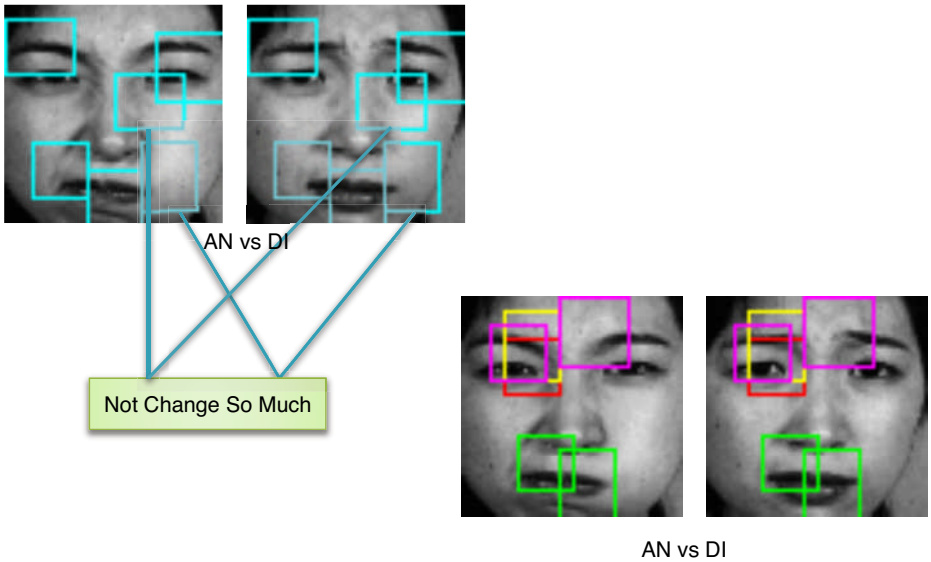


Fig. 4.36: Comparison of AN vs DI.

in AN and DI facial expression are treated as input, and the average accuracy is recorded. From these results, it can be seen that boosted local patterns can keep more discriminant information but with a significant low time cost advantage. And Boosted-1DLPMS-B can achieve the best performance.

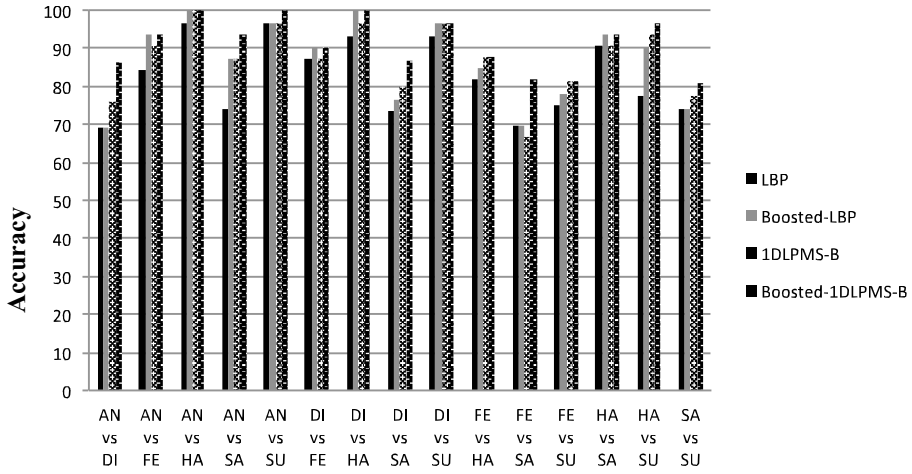


Fig. 4.37: Accuracy (%) of facial expression recognition according to one against one strategy.

4.8 Conclusions

In this chapter, firstly, in order to consider the spatial structure information of our facial images, two variants of L1 norm based LDA are proposed. Next, one effective classifier called Weighted Histogram Spatially constrained Earth Mover's Distance (WHSEMD) that uses the discriminative powers of various image regions, the different patterns and the different spatial information of images, is studied. 2DLDA-L1 and BLDA-L1 are not only robust to outliers or occlusion, but also treat an image as a matrix to make good use of the spatial structure information. BLDA-L1 is a general framework of LDA-L1 and 2DLDA-L1. Our proposed L1-norm optimization technique avoids to compute the eigen-decomposition problem. In detail, they first suppress the negative effects of outliers or occlusion, second they have no 3S problem and third they are invariant to rotations.

Based on the multi-scans based invariant feature descriptors described in the previous two chapters and our above proposed feature projection and classification approaches, huge experiments including face identification, gender estimation and facial expression recognition are evaluated. Experimental results have demonstrated the effectiveness of our proposed methods compared to the existing approaches, especially in the facial occlusion case.

5

Texture Classification with Multi-Scans based Invariant Feature Descriptors

5.1 Introduction

During lots of features, texture is treated as one of the most key and powerful characteristics since it can be extracted from almost every subject in the world. No matter in academia or industry, texture plays a significant position in the recognition, classification or identification systems. Thus, how to get a robust and efficient texture classification approach is one of the hottest topics for researchers. In general, texture is defined as the relationship between the neighborhoods in the surface of the object. And there are two main steps for texture classification: learning stage and recognition stage. In the learning stage, some features should be extracted from the gallery with the known labels to construct a feature database, while the unknown sample is assigned into a label by comparing the extracted features from the probe and the feature database in the classification step. Therefore, the main purpose of texture classification is that how to obtain a best-matched label f

.

A number of texture classification methods have been proposed in the recent decades. In ref. [70] Porter et al. proposed Rotation invariant Daubechies wavelets transform features (RDBWP) to deal with this problem. It is an extension version of his previous work with the advantage of rotation invariant property. In this approach, in order to get rotation invariant features, firstly, the input sample should be decomposed into three levels, and then at each level the mean value of the corresponding averaged norms of the coefficients for some channels is extracted. In Ref. [71], Ojala et al. proposed a novel texture classification algorithm by applying Local Binary Patterns, which is proved to be one of the most effective but simple descriptors to deal with the local features in the image surface. For example, in there researches, the recognition rate of face recognition and texture classification is comparable. In 2005, a novel approach called MR8, was proposed by Varma and Zisserman [72] [73]. The key idea of this method is that it can extract a rotation robust texton library from the gallery and then an unknown probe texture is labeled based on the texton distribution. The common advantage of both LBP and MR8 is that they are statistic approaches but can yield powerful classification performances on challenging and large datasets. Generally, these approaches can be summarized into three stages: feature extraction, histogram generation, and classification. Especially, at the stage of feature extraction, LBP applies a series of linear filters or nonlinear filter [74] to construct some pre-defined patterns, and then these filters are analyzed to judge the contrast at every pixel, while MR8 uses thirty-eight linear filters for extracting an eight dimension feature space vector at every pixel. Thus, MR8 is time consuming and LBP is sensitive to noise and non-monotonic illumination change. A sparse texture representation was proposed by Lazebnik et al. in Ref. [75]. In this method, a wide range of transformations, such as non-rigid deformation and various viewpoint, are applied to extract local affine regions to recognize the surfaces of textures. Earth Movers Distance (EMD) is applied to analyze the similarity of

these two signatures. The disadvantage of it is that in the learning set, auxiliary labels should be included to specify the groups of the images. In ref. [76], Mellor et al. presented a novel approach based on the invariant combinations of linear filters. Different from Lazebniks algorithm, a new set of filters is designed that can provide scale robust features, as a result, texture descriptors which are invariant to local changes at scale, contrast, orientation and less sensitive to local skew can be got. Recently, In Ref. [77], a new unsupervised method for texture classification was proposed by Qin et al. The key idea is that they represent the texture sample as a feature vector by a bag-of-keypoints, and the keypoints are extracted from a set of invariant descriptors for each sample. In ref. [78], SURF feature was applied to do texture classification.

In the classification stage, various classification methods have been explored from nearest-neighbor matching [73] to kernel-based SVMs [79]. The superiority of SVMs for texture classification has been clearly demonstrated, but nearest-neighbor is still often used as an uncommitted mechanism to compare texture representations due to its simplicity and absence of parameters that need to be tuned.

In the chapter 4, L1-norm based LDA and its variants have show effectiveness to deal with outliers or occlusion problems, but if the percent of outliers or occlusion is very large, such as more than 40% compared to useful information, these approaches cannot obtain satisfied performance. Then in this chapter, by replacing the maximize variance that is based on L1-norm or L2-norm, maximum correntropy criterion (MCC) [80] based Linear Discriminant Analysis (we denote it as LDA-MCC) is proposed, which is a powerful measurement to deal with non-Gaussian noise with large outliers. Based on the point of Information Theoretic Learning (ITL), LDA-MCC is an improved extension version of LDA by replacing MSE principle with MCC but with some appealing merits: 1) LDA-MCC is more robust to large outliers as well as it is rotationally invariant. 2) The optimal solutions for our proposed approach

are the principal eigenvectors that are corresponded with the largest eigenvalues in a robust covariance matrix .

In general, Linear Discriminant Analysis (LDA) based on a new Maximum Correntropy Criterion optimization technique is capable of decreasing the influence of outliers significantly, especially, in the case with large outliers, resulting in a robust classification and can be effectively solved by half-quadratic optimization algorithm.

In this chapter, we will test our proposed multi-scans based invariant feature descriptors and classifiers into texture classification based on two famous and challenging databases - Brodatz [81], and CUReT [82].

The remainder of this chapter is organized as follows: Maximum Correntropy Criterion based Linear Discriminant Analysis (LDA-MCC) will be introduced in section 2. Experiments based on texture classification will be reported in section 3 and 4. Finally, conclusions are discussed in section 5.

5.2 Maximum Correntropy Criterion based Linear Discriminant Analysis

5.2.1 Problem Formulation

Assume we have a set of samples $X = \{\{x_i^l\}_{i=1}^{N_l}\}_{l=1}^{C} \in \mathbb{R}^{d\times n}$, N_l of which belong to class ω_l $(l = 1, 2, ..., C)$, where n is the number of samples while d is the dimension of the original input space. And $n = \sum_{l=1}^{C} N_l$. In LDA-L2, the objective is to seek t projections $Y = \{\{y_i^l\}_{i=1}^{N_l}\}_{l=1}^{C} \in \mathbb{R}^{t\times n}$ by means of t linear transformation vectors $W \in \mathbb{R}^{d\times t}$, which embeds the original d dimension into t dimension vector space such that $t < d$. Let Tr(.) denote the trace of the matrix argument, S_b denote the between-class scatter matrix, and S_w denote the within-class scatter matrix, which are formulated as: $S_b = \sum_{l=1}^{C}(m_l - m)(m_l - m)^T$ and $S_w = \sum_{l=1}^{C}\sum_{i=1}^{N_l}(x_i^l - m_l)(x_i^l - m_l)^T$. Here $m_l = (1/N_l)\sum_{i=1}^{N_l} x_i^l$ is the mean of the samples that are assigned to class ω_l, and $m = (1/n)\sum_{l=1}^{C} N_l m_l$ is the global mean of all the samples. The purpose of LDA-L2

is to obtain an optimal transformation W by maximizing the ratio of $\mathrm{Tr}(S_b)$ and $\mathrm{Tr}(S_w)$ as following problem

$$\begin{aligned} \max_W J_{L2} &= \max_W \frac{\mathrm{Tr}(S_b)}{\mathrm{Tr}(S_w)} \\ &= \frac{W^T S_b W}{W^T S_w W}. \end{aligned} \tag{5.1}$$

The denominator of the objective function J_{L2} can be simply to $W^T S_w W = I$, since it is invariant with respect to rescaling of the vectors $W \to \beta W$ (β is some coefficient). Thus, the problem of maximizing J_{L2} can be converted into the following constrained optimization problem:

$$\begin{aligned} \max_W \quad & W^T S_b W, \\ s.t. \quad & W^T S_w W = I. \end{aligned} \tag{5.2}$$

It is well known that the L2-norm is not robust to outliers and recently, R1-norm approach [23] was presented to solve this problem. Thus, the issue can be defined to find W which maximizes the below objective function:

$$\begin{aligned} \max_W J_{R_1} = (1-\alpha) \textstyle\sum_{l=1}^{C} \sqrt{||W^T(m_l - m)||^2} - \\ \alpha \textstyle\sum_{l=1}^{C} \sum_{i=1}^{N_l} \sqrt{||W^T(x_i^l - m_l)||^2}. \end{aligned} \tag{5.3}$$

Here, α is a predefined parameter to get the trade-off between first term and second term. However, it takes lots of cost to obtain final convergence in a large input dimension space, and in the spatial dimension, squared data is still used. Thus, R1-norm approach is not effective and efficient for larger outliers problems. In this chapter, we try to use Maximum Correntropy Criterion (MCC) to measure the between-class scatter instead of Mean Square Error (MSE). In practice, the definition of correntropy is a generalized similarity measure distance between two arbitrary random variables A and B:

$$V_{n,\sigma}(A,B) = \frac{1}{n}\sum_{l=1}^{n} k_\sigma(a_l - b_l). \tag{5.4}$$

When kernel f k (.) is Gaussian kernel $g(x) = exp(-x^2/2\sigma^2)$, then

$$V_{n,\sigma}(A,B) = \frac{1}{n}\sum_{l=1}^{n} g(a_l - b_l). \tag{5.5}$$

In order to measure the similarity of two random variables A and B, MSE uses all the samples in the input space while correntropy is just determined by kernel f a_l b_l. This important property intuitively explains the reason that the correntropy is superior to MSE if the residual of $A-B$ is non-symmetric or with nonzero mean.

In ITL, it has been pointed out that MSE is a global measurement while MCC is a local measurement [80]. That means, in MSE measurement, the contribution of each data in the joint space is equal to the final decision while the major judgment in MCC scheme is determined by the kernel f lated points. In general, the outlier point is far from the data category, so its contribution to the final correntropy decision is weak, and a low rank value in the matrix is returned. Therefore, in MCC scheme, with the increasing of correntropy, the influence of outlier will be decreased on the final decision. As a result, LDA-MCC is robust against outliers even large outliers occur.

Substituting $a_l = (m_l - m)$ and $b_l = WV_l$ into Eq.(5.5), here, $V_l = W^T(m_l - m)$ is a projected vector, and we can obtain a novel maximum correntropy criterion based LDA as follows:

$$\begin{array}{ll} \underset{W}{max} & J_{MCC} = \sum_{l=1}^{C} g((m_l - m) - WV_l), \\ s.t. & W^T S_w W = I. \end{array} \tag{5.6}$$

Since W is orthonormal and then

$$\begin{aligned} &g((m_l - m) - WV_l) = g(\sqrt{||(m_l - m) - WW^T(m_l - m)||^2}) \\ &= g(\sqrt{(m_l - m)^T(m_l - m) - (m_l - m)^T WW^T(m_l - m)}). \end{aligned} \tag{5.7}$$

Let $M_l = (m_l - m)$, then the Eq.(5.6) can be converted into following objective function:

$$\begin{aligned} \underset{W}{max} \quad & J_{MCC} = \sum_{l=1}^{C} g(\sqrt{M_l^T M_l - M_l^T WW^T M_l}), \\ s.t. \quad & W^T S_w W = I. \end{aligned} \tag{5.8}$$

5.2.2 Algorithm

Recently, ITL has been proved more efficient to data analysis problems. ITL utilizes probability density function of the data, estimated by Parzen kernel estimator [83], as the cost function.

5.2.3 Optimization

In literature, in order to optimize nonlinear ITL issue, the half-quadratic approach [84] [85] is always imported. And also in this thesis, half quadratic based method is used to solve Eq.(5.8). Based on the principle of convex conjugated functions [84], the following proposition can be obtained.

Proposition: There has a convex conjugated function φ of $g(x)$ like

$$g(x) = \underset{p'}{max} \left(p' \frac{||x||^2}{\sigma^2} - \varphi(p') \right), \tag{5.9}$$

where $p' \in R$ is a scalar parameter, and the maximum is obtained at $p' = -g(x)$ for a fixed x [85]. Substituting Eq.(5.9) into Eq.(5.8), an augmented objective equation in a enlarged parameter dimension is obtained, then the Eq.(5.6) can be converted into

$$\begin{array}{ll} \underset{W,P}{max} & J_{MCC} = \sum_{l=1}^{C}(p_l(M_l^T M_l - M_l^T W W^T M_l) - \varphi(p_l)), \\ s.t. & W^T S_w W = I, \end{array} \tag{5.10}$$

where $P = [p_1, p_2, ..., p_C]$ is storing the auxiliary variables imported to solve the Half-Quadratic issue. Consequently, we can optimize (W, P) by iterations as:

$$\underset{W,P}{max}\, \mathcal{L} = J_{MCC} - \lambda(W^T S_w W - I). \tag{5.11}$$

Then, according to Lagrangian method, a weighted traditional LDA problem can be obtained as follows

$$(S_w)^{-1} S_b P W = \lambda W, \tag{5.12}$$

where P is a diagonal matrix with diagonal entity $p(l, l) = -p_l$ and

$$p_l = -g(\sqrt{M_l^T M_l - M_l^T W W^T M_l}). \tag{5.13}$$

Thus, the final algorithm of LDA-MCC is listed in Algorithm 2.

Algorithm 2 LDA-MCC

Require: $X = \{\{x_i^l\}_{i=1}^{N_l}\}_{l=1}^{C} \in \mathbb{R}^{d \times n}, t \leq d.$
Initialization: $W = [w_1, w_2, ..., w_t] \in \mathbb{R}^{d \times t}, W^T W = I.$
while not converge **do**
1. Calculate $p_l = -g(\sqrt{M_l^T M_l - M_l^T W W^T M_l})$.
2. Update W according to $(S_w)^{-1} S_b P W = \lambda W$.
end while
return $W \in \mathbb{R}^{d \times t}$.

5.2.4 Convergence of LDA-MCC

Let r be the iteration number of Algorithm 2. then

$$\begin{array}{l} J_{MCC}^{r+1} - J_{MCC}^{r} = J_{MCC}(W^{r+1}, P^{r+1}) - J_{MCC}(W^r, P^r) \\ \quad = [J_{MCC}(W^{r+1}, P^{r+1}) - J_{MCC}(W^r, P^{r+1})] \\ \qquad + [J_{MCC}(W^r, P^{r+1}) - J_{MCC}(W^r, P^r)]. \end{array} \tag{5.14}$$

Based on the Proposition and Eq.(5.12), W^{r+1} and P^{r+1} is the optimization value for J^{r+1}_{MCC} and J^{r}_{MCC}, respectively. Then $J_{MCC}(W^{r+1}, P^{r+1}) - J_{MCC}(W^{r}, P^{r+1}) \geq 0$ and $J_{MCC}(W^{r}, P^{r+1}) - J_{MCC}(W^{r}, P^{r}) \geq 0$. So $J^{r+1}_{MCC} - J^{r}_{MCC} \geq 0$. That is, the objective function $J^{r}_{MCC}|_{r=1,2,...}$ increases monotonically. In the other side, apparently, $J^{r}_{MCC}|_{r=1,2,...}$ function has an upper bound. Thus, we can get that $J^{r}_{MCC}|_{r=1,2,...}$ converges.

5.2.5 Evaluation on Toy Sets

This subsection is to evaluate our proposed LDA methods in a simple but effective way, which is based on a toy set consisted of ten samples that are clustered to two category with an additional small and large outliers as shown in Fig.5.1(a) and Fig.5.2(a), respectively.

To evaluate the effectiveness of LDA-L1 and LDA-MCC which is less sensitive to outliers, the outlier sample 1, which is illustrated as triangle at the top-right corner of Fig.5.1(a), is intentionally added into the training data in Class one before classification. For this kind of data, LDA-L2, LDA-R1, LDA-L1 and LDA-MCC are applied and the projection vectors are $w_{L2} = [-0.7071, 0.7071]^T$, $w_{R1} = [-0.76431, 0.6448]^T$, $w_{L1} = [-0.8784, 0.4779]^T$ and $w_{MCC} = [-0.8784, 0.4779]^T$, respectively. The final learned results are describted as 1-dimensional signals in Fig.5.1(b) , Fig.5.1(c), Fig.5.1(d) and Fig.5.1(e) corresponding to LDA-L2, LDA-R1, LDA-L1 and LDA-MCC, respectively. From these figures ,we can see clearly that after the process of dimension reduction, the between-class scatters of the two-category data excluded the outlier data in Fig.5.1(d) and Fig.5.1(e) are much larger than that in Fig.5.1(b) and Fig.5.1(c). In this experiment, LDA-MCC is initialized randomly and just two iterations are applied for convergence, while LDA-R1 converges in four iterations. Thus, our proposed methods are more powerful to address the outliers' problem.

In order to show how the LDA-MCC can deal with larger outliers problems,

the outlier sample 2, which is illustrated as triangle at the top-right corner of Fig.5.2(a), is intentionally added into the training data in Class one for learning. For this kind of data, LDA-L2, LDA-R1, LDA-L1 and LDA-MCC are applied and the projection vectors are $w_{L2} = [-0.7071, 0.7071]^T$, $w_{R1} = [-0.71532, 0.69879]^T$, $w_{L1} = [-0.71953, 0.69447]^T$ and $w_{MCC} = [-0.8784, 0.4779]^T$, respectively. The final learned results are describted as 1-dimensional signals in Fig.5.2(b) , Fig.5.2(c), Fig.5.2(d) and Fig.5.2(e) corresponding to LDA-L2, LDA-R1, LDA-L1 and LDA-MCC, respectively. From these figures, we can see clearly that even the outliers are very large, the LDA-MCC can also get the stable performance, while LDA-R1 and LDA-L1 can not deal with this issue effectively.

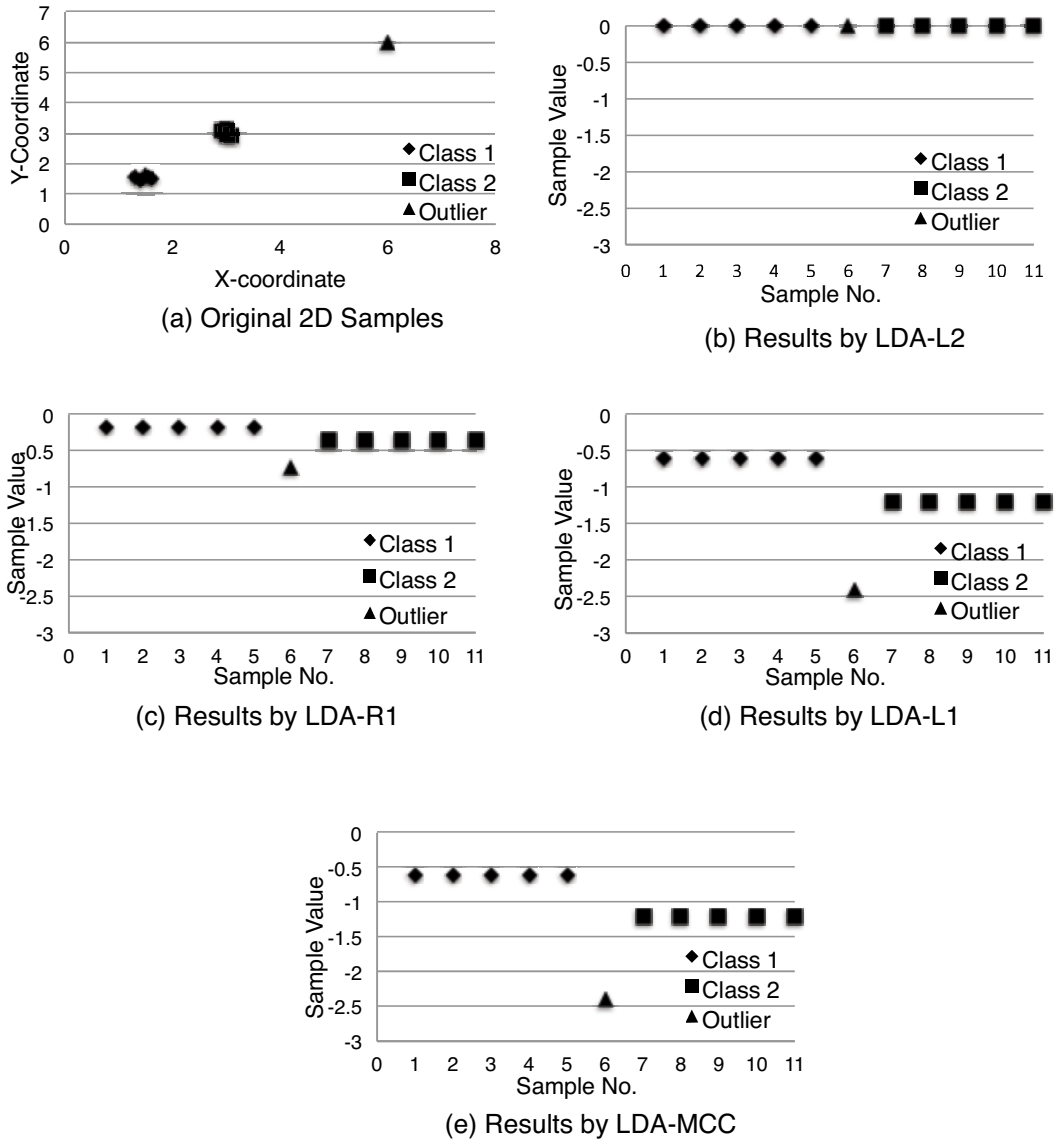
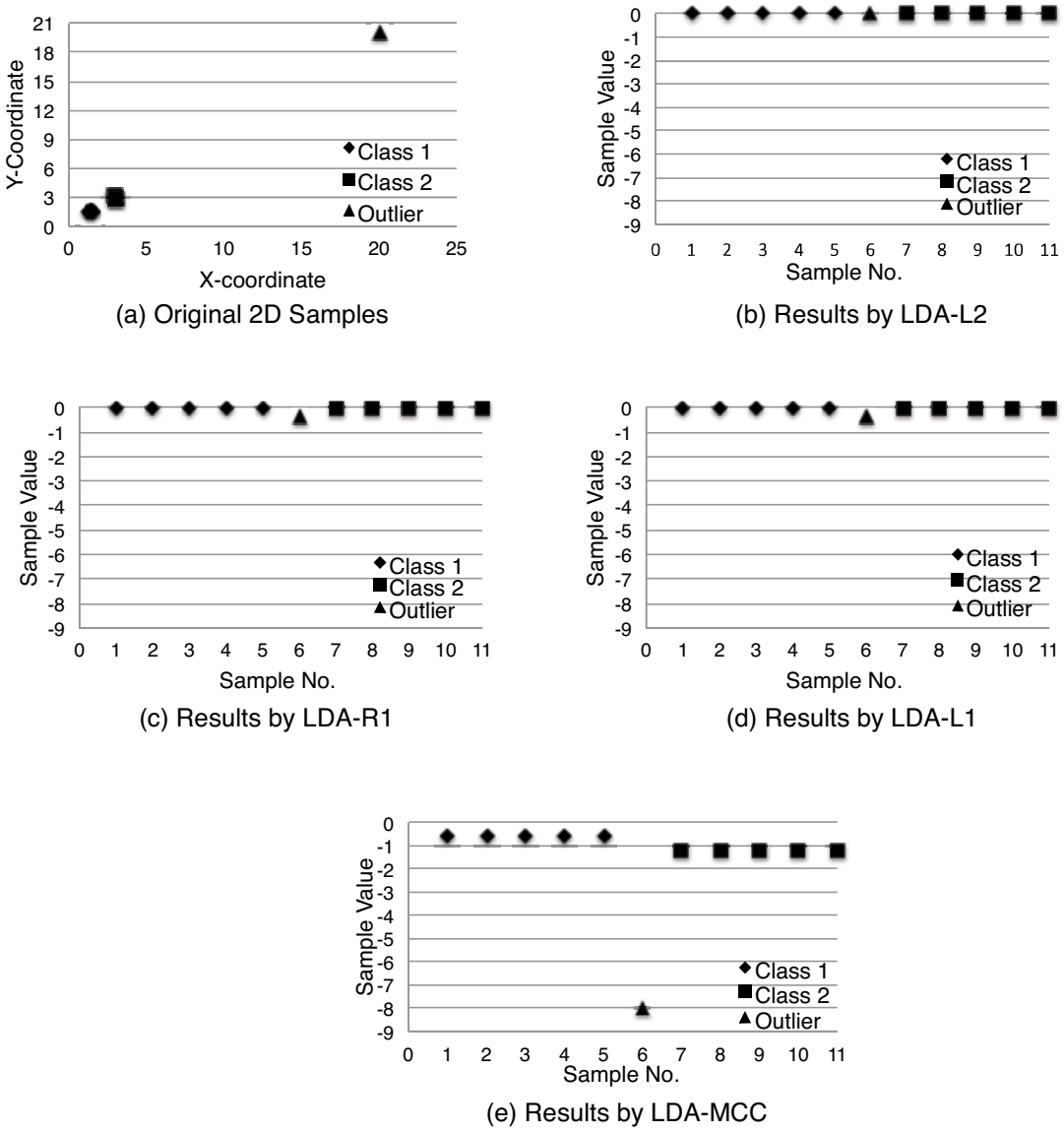


Fig. 5.1: Results obtained in small outliers case. (a) Samples in toy set 1 with small outliers. (b) Results of LDA-L2. (c) Results of LDA-R1. (d) Results of LDA-L1. (e) Results of LDA-MCC.



(a) Original 2D Samples

(b) Results by LDA-L2

(c) Results by LDA-R1

(d) Results by LDA-L1

(e) Results by LDA-MCC

Fig. 5.2: Results obtained in large outliers case. (a) Samples in toy set 2 with large outliers. (b) Results of LDA-L2. (c) Results of LDA-R1. (d) Results of LDA-L1. (e) Results of LDA-MCC.

5.3 Brodatz Texture Database

In this section, an old but well-known texture dataset called Brodatz dataset is evaluated. This dataset is developed from the Brodatz Album [86] that includes 111 samples, and it is constructed by separating the every picture of Brodatz Album into 9 non-overlapping 215 $\times$ 215 samples. So totally there are 999 samples on the Brodatz database. Since the number of group on the Brodatz dataset is very big with a relatively small number of samples are included per group, it is more challenging to deal with texture classification issue. Fig.5.3 shows some samples of the four categories D35, D64, D74 and D99 in this database.

In the first evaluation, the mean classification rate for all textures is estimated. In this experimental setup, 3 images are used for training and others are for testing. The final results are concluded in Table. 5.1. From this table, we can conclude that our proposed multi-scans based invariant feature descriptors are powerful to tackle texture classification, while LLCP is the outstanding one since LLCP both considered the global texture and local texture information. In addition, some texture-specific codebooks are learned for more accurate representation.

Table 5.1: Accuracy on Brodatz database.

Method	Accuracy (%)
Lazebnik et al. [75]	88.15
Mellor et al. [76]	89.71
Qin et al. [77]	64.46
Soottitantawat et al. [78]	90.84
LBP	83.12
1DLPMS-B	89.82
LCBPMS	93.96
LLCP	97.65

The second experiment is to evaluate the classification performance over the subset of Brodatz Texture Dataset [81] with the variations of rotation. In this evaluation, 13 images are selected as categories (shown in Fig.5.4). This sub-database is a ro-

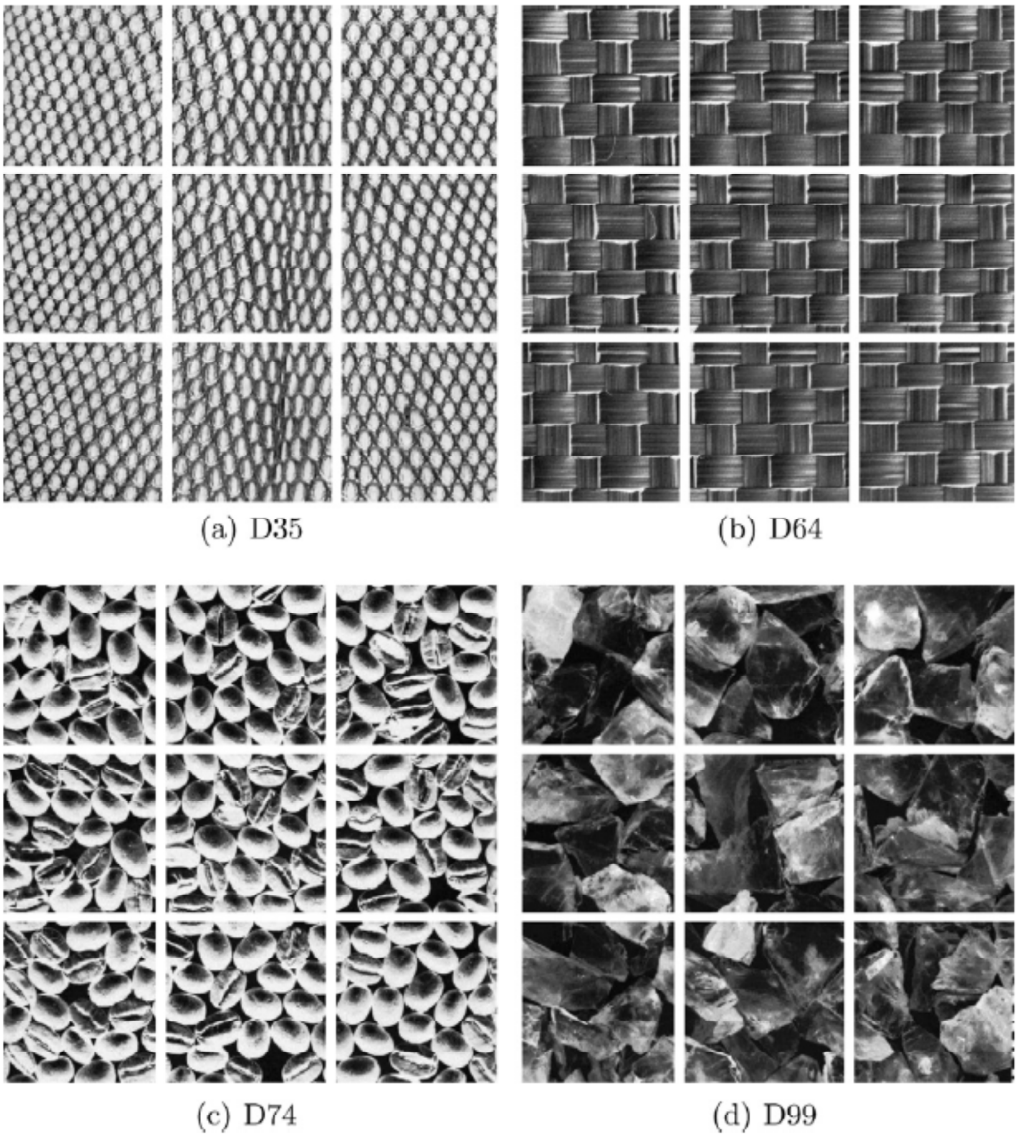


Fig. 5.3: Samples on Brodatz texture database.

tated version of Brodatz database and consists of thirteen images each digitized at seven different rotation directions: 0, 30, 60, 90, 120, 150, and 200 degrees. Each image is normalized into 128 × 128 size, and then is non-overlapping divided into 16 regions. The images are all 32 × 32 pixels with 8 bits/pixel. 12 images of per class are used for gallery and others are probe.

The precisions of some traditional approaches and our proposed methods are list

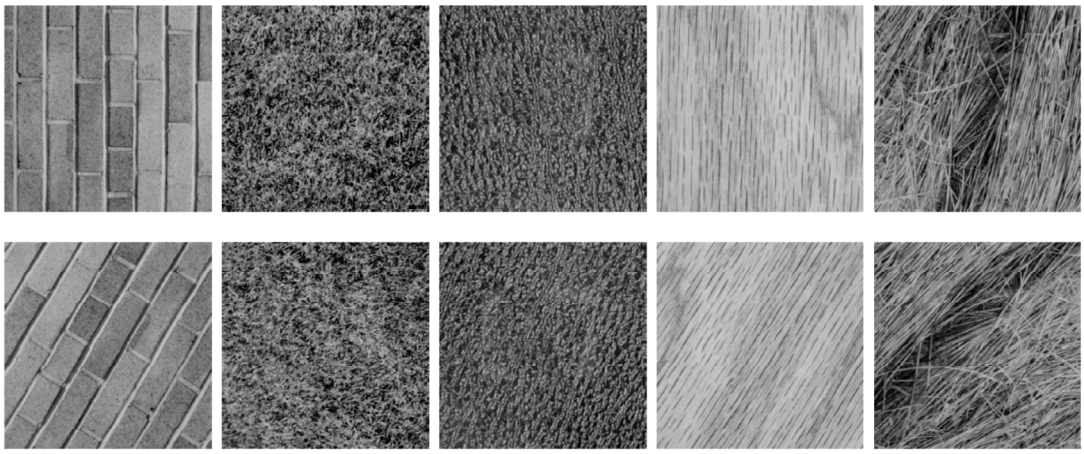


Fig. 5.4: Samples on Brodatz texture database with rotated ones. The first row shows five reference images: brick (D94), grass (D9), leather (D24), wood (D68), and straw (D15), the second row shows the rotated versions of the reference images.

in Table 5.2. From this table, we can also see the effectiveness of our proposed methods with the variation of rotations in texture classification application.

Table 5.2: Accuracy on Brodatz database with variation of rotations.

Method	Accuracy (%)
LBP	73.3
1DLPMS-B	89.3
1DLPMS-T	92.5
weighted LCBPMS	94.9
weighted LLCP	96.2

The third experiment is to evaluate the classification performance over the subset of Brodatz Texture Dataset [81] to test the outliers effect. In the first estimation, 20 images are selected as categories("real" images) and one image is selected as outliers image(shown in Fig.5.5).

At first, each image is normalized into 128×128 size, and then is non-overlapping divided into 16 regions. 5 regions per category and 1 region in outliers image are used as gallery and others per category are treated as probe. The final classification performance is shown in Fig.5.6, where x-axis is associated with the reduced dimension

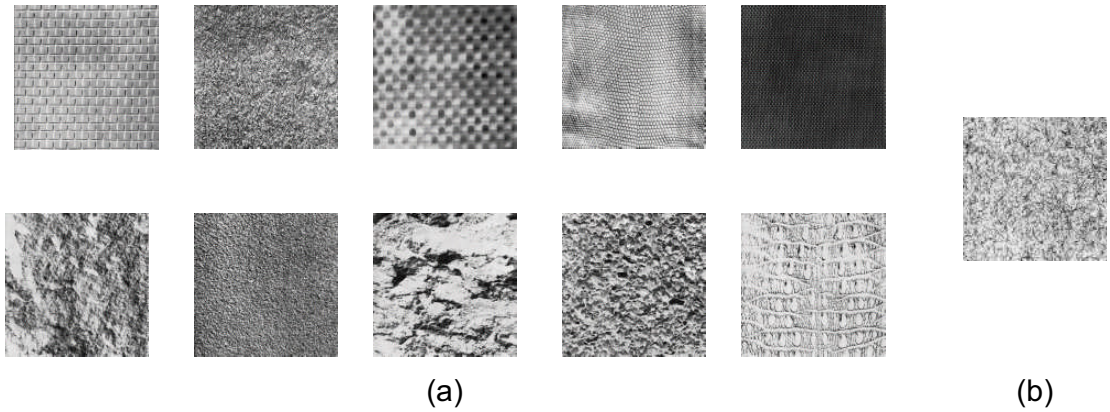


Fig. 5.5: Samples on Brodatz texture database. (a) "Real" images. (b) Outliers image.

and y-axis corresponds to the accuracy. From this figure, we can see our proposed method is less sensitive to outliers than the other two traditional approaches. In average, our proposed methods can achieve about 6 percent higher than LDA-R1 and 11 percent higher than LDA-L2 approach. Moreover, form Dim. 30 to Dim. 40, the accuracy of LDA-R1 drops significantly, that means the projection weights from Dim. 30 to Dim. 40 obtained by LDA-R1 are very sensitivity to the outlier while LDA-MCC is much stable and a little better than LDA-L1. The final recognition result with our proposed feature extraction methods and some traditional methods are summarized in Table. 5.3. From this table, we can clearly see the effectiveness of our proposed approaches.

In order to show how the classification methods affect when large outliers contained, in the next experiment, different from the previous evaluation, 20 images are selected as categories ("real" images) but four images are selected as outliers image. Then 5 regions per category and 4 regions in outliers image are used as gallery and others per category are treated as probe. The final classification results are shown in Fig.5.7. Compared with Fig.5.6, we can see that the performance of LDA-L2, LDA-R1 and LDA-L1 drop about 15 percent, 10 percent and 8 percent, respectively,

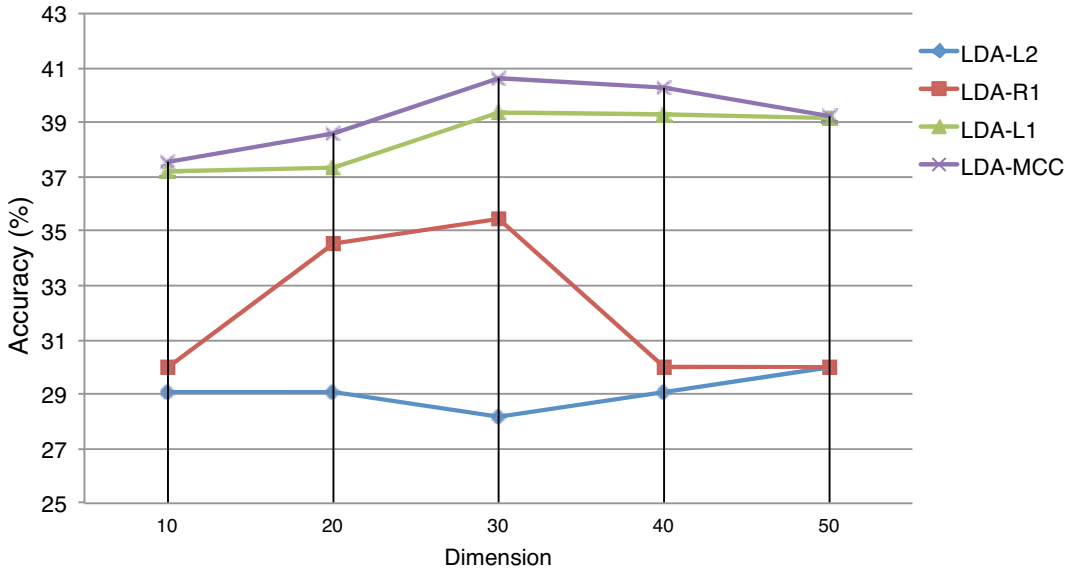


Fig. 5.6: Accuracy on Brodatz texture database with small outliers.

Table 5.3: Accuracy on Brodatz database with small outliers.

Method	Accuracy (%)
LDA-L2	30.0
LDA-R1	35.4
LDA-L1	39.4
LDA-MCC	40.6
LBP	67.5
LGBP+LDA-R1	84.2
1DLPMS-T	83.3
weighted LCBPMS	86.3
weighted LLCP	91.5
LCBPMS + LDA-L1	88.8
LLCP + LDA-L1	92.3
LCBPMS + LDA-MCC	89.8
LLCP + LDA-MCC	93.1

while LDA-MCC can get almost stable result.

Also, the final classification result with our proposed feature extraction methods and some traditional methods are summarized in Table. 5.4. From this table, we can conclude that the traditional LBP is sensitive to the outliers problems when our proposed feature extraction methods are much more robust and stable.

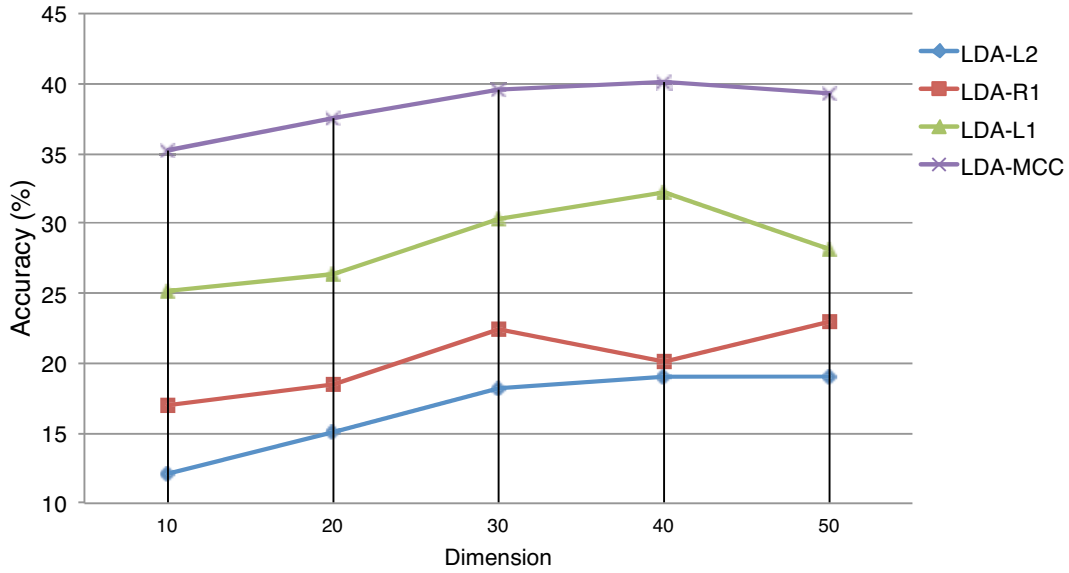


Fig. 5.7: Accuracy on Brodatz texture database with large outliers.

Table 5.4: Accuracy on Brodatz database with large outliers.

Method	Accuracy (%)
LDA-L2	19.1
LDA-R1	23.0
LDA-L1	32.2
LDA-MCC	40.1
LBP	45.6
LGBP+LDA-R1	77.2
1DLPMS-T	80.1
weighted LCBPMS	84.2
weighted LLCP	89.3
LCBPMS + LDA-L1	85.1
LLCP + LDA-L1	89.5
LCBPMS + LDA-MCC	88.9
LLCP + LDA-MCC	92.1

5.4 CUReT Database

The Columbia-Utrecht Reflectance and Texture database (CUReT) [82] contains images of 61 different real world textures each imaged under more than 200 different viewpoints and under different illuminations (www1.cs.columbia.edu/CAVE/ software/curet/). Some examples of 61 CUReT textures are shown in Fig. 5.8. There

are 205 samples of each texture captured with the variations of various illumination directions and viewpoints, while there are also 118 samples with different viewing angles less than 60^o. In our study, among these 118 samples, 92 samples are selected based on the principle that all these texture classes can be cropped into a large region, such as 200 $\times$ 200. In final, all the samples are converted into gray-scale image [75]. Since there are many variations with different viewing angles, various orientations and varied lighting conditions in this dataset, which makes this dataset more challenging but widely studied.

In this experiment, two different setups are performed as shown in the following:

1. T46 where the samples in the gallery set and the probe set are same. That means, in the both training and testing stages, there are 2806 (61 $\times$ 46) samples are used, respectively.

2. T23 where the samples in the gallery set and probe set are different. The ratio between galley and probe is 1:3. That is, there are 403 (61 $\times$ 23) samples and 4209 (61 $\times$69) samples for learning and testing, respectively.

T46 can be seen if we have enough learning data while T23 is treated as the condition that only partial training data can be got. The final classification performances are listed in Table. 5.5. "weighted LLCP" approach is the most excellent one since it both considered the global and local information in the textures and weighting strategy is effective to do matching.

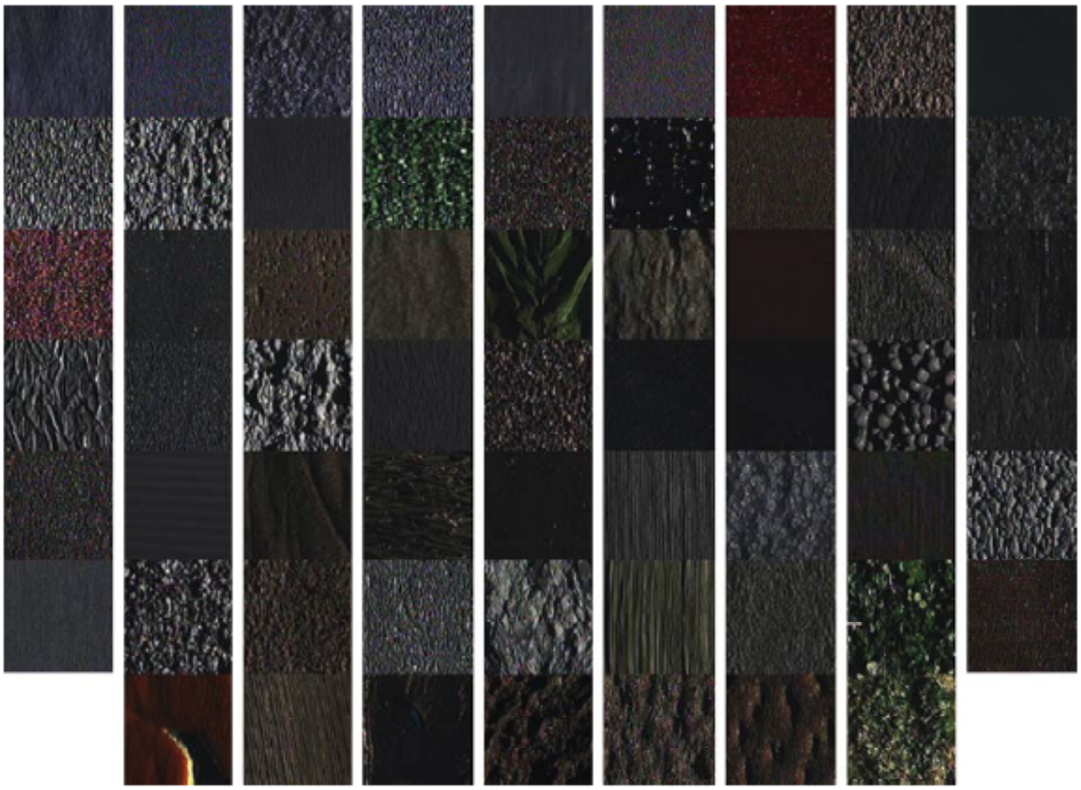


Fig. 5.8: Samples on CUReT Texture database.

Table 5.5: Accuracy (%) on CUReT database.

Method	T46	T23
VAR [87]	68.6	60.9
LBP	81.8	57.9
LTP	85.1	79.2
LBPV [87]	93.1	75.2
MR8 [73]	97.5	77.5
1DLPMS-T	94.6	82.4
weighted LCBPMS	96.3	83.6
weighted LLCP	98.2	89.5

5.5 Conclusions

In this chapter, firstly, we propose a novel method of LDA with MCC, which further better characterizes the between-class separability. Our proposed objective target

is less sensitive to large outliers and is efficiently solved by the half-quadratic optimization method. For every iteration step, this complex correntropy function is able to reduce to a weighted traditional LDA optimization issue. Our proposed subspace approach powerfully suppresses the negative effects of the outliers with the property of invariant to rotation. Next, two large and challenging texture databases are evaluated. Brodatz database is focus on the variations of scale, rotation and outliers while CUReT database is focus on the variations of rotation, illumination and viewpoints. From the experimental results, we can recognize that our proposed feature extraction methods are invariant to changes of outside factors when our proposed feature projection methods can tackle the large outliers problem effectively.

6

Conclusions and Future Work

6.1 Conclusions

Significant researches are done in this work with an aim of designing a reliable automated recognition system which is robust under varying conditions of scale level, rotation, illumination, outliers, occlusion and so on. For specially, our proposed multi-scans encoding model can get invariant feature descriptors under the variation of scale, rotation illumination etc. And our new feature projection strategy is capable of reducing the effect of outliers or occlusion effectively and efficiently to capture the key instinct structure of the data which is located in high dimension for robust classification.

Some of the key points in this dissertation are summarized int the following:

• Several novel invariant feature descriptors based on multi-scans encoding model are proposed in chapter 2. These feature descriptors are scale, rotation and illumination invariant and more robust to other outside factor changes. Compared to traditional approaches, several aspects are improved: the first one is the improvement of the discriminative capability; Second is the enhancement of their robustness;

Third is the selection the neighborhood and fourth is reducing the size of local patterns. For example, ternary and quaternion based multi-scans encoding model can get more discriminative patterns as well as can enhance the robustness of feature descriptor. Experiments on a general database have been verified the effectiveness of our proposed feature descriptors.

- In chapter 3, we focus on further improving the performance of the feature descriptors under different variations, especially, facial expression and pose issues in face recognition application. Then, the multi-scans encoding model is extended into Curvelet frequency space. Curvelet based feature descriptors are a better choice compared with wavelet and Gabor wavelet based ones, since the Curvelet transform has the property to represent the images more sparsely. The most advantage of it is it could extract high degree of orientation, anisotropy and time frequency resolution, which can generate more powerful and effective descriptors in images or textures, such as facial samples. Several pattern-specific codebooks are generated from the training set, which is more suitable for pattern perception tasks. multi-mapping strategy is also studied to improve the precision. In order to select more compact and discriminative features and solve the outliers or occlusion problem as well as keeping intrinsic and discriminative structure from high dimensional data, L1-norm based LDA is proposed and shows very powerful in this aspect. Our proposed feature projection strategy better characterizes the between-class separability and within-class compactness and has many interesting points. For example, our proposed L1-norm optimization technique avoids to compute the eigen-decomposition problem and is easy to implement. Specifically, our proposed feature projection method first suppresses the negative effect of outliers or occlusion, secondly it has no 3S problem since it does not want to compute the inverse of within-class matrix and thirdly it is invariant to rotations. Evaluation based on some databases illustrates the powerful of our proposed approaches.

• Chapter 4 has focused on the face evaluation with our proposed approaches. Firstly, in order to take the spatial structure information of our face into account, the LDA-L1 proposed in chapter 3 is extended from 1D to 2D and two variants called 2DLDA-L1 and BLDA-L1 are proposed, which are not only robust to outliers and occlusion, but also treat an image as a matrix to make good use of the spatial structure information. Secondly, an classification method called Weighted Histogram Spatially constrained Earth Mover's Distance (WHSEMD), which illustrates the discriminative powers of various image regions, the different patterns and the different spatial information of images, is proposed. Experiments based on face recognition problem including face identification, gender estimation and facial expression evaluation show the effectiveness and efficiency of our proposed algorithms not only on occluded case but also on un-occluded case. From the experiments, we can see that the accuracy of face identification is improved about 10%-30% while time cost is reduced about 1/4 by our proposed feature extraction methods in chapter 2. Multi-scans based invariant feature descriptors are powerful to select the discriminative feature in our facial images by adaboost technology. Experiments based on famous and challenging face databases - FERET [1], LFW [55], FRGC [56] show that our proposed methods in chapter 3 can further improve 10%-20% recognition rate compared to our proposed approaches in chapter 2. Our proposed feature projection methods can solve the outliers or occlusion problems significantly. According to the gender estimation, we can conclude that male is expected easier to be recognized in our real life while comparable with female if the frontal face is used.

• In order to further reduce the effect of larger outliers, Linear Discriminant Analysis (LDA) based on Maximum Correntropy Criterion(MCC) optimization technique is proposed in chapter 5, which further better characterizes the between-class separability. The merit is greatly obtained by the kernel measurement in MCC strategy. Since an outlier is located far away to the data category, then its judgement to es-

timate the correntropy is weak with a small rank value in the final decision. Thus, the outliers will give smaller influence on the decision with the increase of correntropy. Our proposed algorithm powerfully suppresses the negative effects of outliers with the property of invariant to rotation as well as it can be efficient and optimal solved by the half-quadratic optimization method. From the experiments based on two famous and challenging databases under the different variations of scale, rotation, outliers and illumination, we can recognize that our proposed approaches can tackle these problems more excellent compared to some state-of-the-art approaches. Especially, even the outliers are very large, our proposed feature projection can deal with them effectively and the accuracy can be improved about 20%.

6.2 Future Work

In this work, the recognition rate of small pose issue, such as images ranging from -30° to +30° pose, can obtain satisfied results. But when the images ranging from -90° to -60° or -60° to -90° pose, this kind of large pose problem cannot be solved significantly. Investigating large pose invariance will lead to a good performance for the pattern recognition system. One possible direction to solve large pose problem is to build three dimensional pattern image. In order to satisfy this condition, our proposed local patterns should be extended to 3D in future. The second future work is that the real time pattern recognition is an aim in many security and surveillance applications. By using hardware to improve the computing speed, we can design a more reliable and fast recognition system [24]. For example, FPGA, ARM, NEON, GPU technologies can be applied to speed up our system significantly. Thirdly, in this study, only gray-scale images are used, and in the future, the color information or other useful direction, such as IR component, could be included to improve the accuracy. With the color information, more useful features can be extracted, while by IR information, recognition system can be worked in the low-lighting environment. Fourthly, only still

images are used in this study, but in many applications, especially in surveillance, video can be recorded, the analysis of the adjacent frames in the video can give us more powerful decision to do recognition tasks. Thus, how to design a system to extract features of video is another direction. Fifth, the feature projection approaches in this work are only considered in 1D and 2D. In order to solve more general pattern recognition problem in high dimensional space, our proposed approaches should be extended to tensor or kernel space, especially for Linear Discriminant Analysis based on Maximum Correntropy Criterion. And since tensor or kernel based methods are usually time computation, multi-linear subspace seems to be one solution. Thus, how to generate multi-linear subspaces from high dimensional data effectively and efficiently is another future work.

Bibliography

[1] P.J. Phillips, H. Moon, P.J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, 2000.

[2] A. Martnez and R. Benavente, "The AR face database," Computer Vision Center (CVC) Technical Report 24, 1998.

[3] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol.60, no.2, pp.91–110, 2004.

[4] H. Ba . Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," Computer Vision and Image Understanding, vol.110, no.3, pp.346–359, 2008.

[5] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.10, pp.1615–1630, 2005.

[6] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.12, pp.2037–2041, 2006.

[7] S. Liao and A. Chung, "Texture classification by using advanced local binary patterns and spatial distribution of dominant patterns," Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1221 – 1224, 2007.

[8] N.S. Vu, H.M. Dee, and A. Caplier, "Face recognition using the POEM descriptor," Pattern Recognition, vol.45, no.7, pp.2478–2488, 2012.

[9] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," IEEE Trans. on Image Processing, vol.19, no.2, pp.533–544, 2010.

[10] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," Proc. of IEEE International Conference on Computer Vision, pp.786 – 791, 2005.

[11] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition," IEEE Trans. Image Processing, vol.16, no.1, pp.57 – 68, 2007.

[12] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, "Learned local gabor patterns for face representation and recognition," Signal Processing, vol.89, no.12, pp.2333 – 2344, 2009.

[13] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, vol.2, pp.506–513, 2004.

[14] M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, pp.586–591, 1991.

[15] B.H. Shekar, M. Sharmila Kumari, L.M. Mestetskiy, and N. Dyshkant, "FLD-SIFT: Class based scale invariant feature transform for accurate classification of faces," Computer Networks and Information Technologies, pp.15–21, 2011.

[16] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," Proc. of International Conference on Automatic Face and Gesture Recognition, pp.336–341, 1998.

[17] C.H. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," Advances in Biometrics, pp.809–818, 2007.

[18] A. Joshi, A. Gangwar, R. Sharma, and Z. Saquib, "Periocular feature extraction based on LBP and DLDA," Advances in Computer Science, Engineering & Applications, pp.1023–1033, 2012.

[19] S. Shan, W. Zhang, Y. Su, X. Chen, and W. Gao, "Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition," Proc. of IEEE International Conference on Pattern Recognition, vol.4, pp.606–609, Aug. 2006.

[20] P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, and G. Aachen, "Surf-face: Face recognition under viewpoint consistency constraints," BMVC, 2009.

[21] N. Kwak, "Principal component analysis based on L1-norm maximization," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.9, pp.1672–1680, 2008.

[22] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2DPCA," IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol.40, no.4, pp.1170–1175, 2009.

[23] X. Li, W. Hu, H. Wang, and Z. Zhang, "Linear discriminant analysis using rotational invariant L1 norm," Neurocomputing, vol.73, no.13, pp.2571–2579, 2010.

[24] A. Nabatchian, "Human face recognition," Doctoral Dissertation, University of Windsor, 2011.

[25] R. Heitmeyer, "Biometric identification promises fast and secure processing of airline passengers," International Civil Aviation Organization's volcanicash contingency plan for the European region (ICAO) Journal, vol.55, no.9, pp.10–11, 27–28, 2000.

[26] W. Zhou, A. Ahrary, and S. Kamata, "Face recognition with local feature patterns and histogram spatially constrained earth mover's distance," Proc. of IEEE International Conference on Signal and Image Processing Applications, pp.374 – 379, 2009.

[27] W. Zhou, A. Ahrary, and S. Kamata, "Face recognition using local quaternion patters and weighted spatially constrained earth movers distance," Proc. of IEEE International Symposium on Consumer Electronics, pp.285 – 289, 2009.

[28] S. Liao, M. w.k.Lawand, and A. Chung, "Dominant local binary patterns for texture classification," IEEE Trans. on Image Processing, vol.18, no.5, pp.1107–1118, 2009.

[29] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms,," Multiscale Modeling and Simulation, vol.5, no.3, pp.861 – 899, 2006.

[30] Available at http://www.curvelet.org.

[31] Available at http://cseweb.ucsd.edu/ naverma/RPTrees/index.html.

[32] J. Landon, B. Jeffs, and K. Warnick, "Model-based subspace projection beamforming for deep interference nulling," IEEE Trans. on Signal Processing, vol.60, no.3, pp.1215 – 1228, 2012.

[33] C.S. Yanlin Geng and P. Hao, "Square loss based regularized LDA for face recognition using image sets," Proc. of IEEE International Conference on Computer Vision and Pattern Recognition Workshops, pp.99–106, 2009.

[34] G.J. McLachlan, Discriminant analysis and statistical pattern recognition, Wiley-Interscience, 2005.

[35] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.7, pp.711–720, 1997.

[36] F. Chen, H.Y. Liao, M.T. Ko, J.C. Lin, and G.J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," Pattern Recognition, vol.33, no.10, pp.1713–1726, 2000.

[37] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," Pattern Recognition, vol.34, no.12, pp.2067–2070, 2001.

[38] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, vol.2, pp.564–569, 2004.

[39] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.2, pp.230–245, 2005.

[40] A.K. Jain, "Biometric recognition: how do I know who you are?," Proc. of IEEE Signal Processing and Communications Applications Conference, pp.3–5, 2004.

[41] S. Zhou, G. Aggarwal, R. Chellappa, and D. Jacobs, "Appearance characterization of linear lambertian objects generalized photometric stereo, and illumination-invariant face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.2, pp.230–245, 2007.

[42] K.C. Kwak and W. Pedrycz, "Face recognition using an enhanced independent component analysis approach," IEEE Trans. Neural Networks, vol.18, no.2, pp.530–541, 2007.

[43] Z.M. Hafed and M.D. Levine, "Face recognition using the discrete cosine transform," International Journal of Computer Vision, vol.43, no.3, pp.167–188, 2001.

[44] J. Yang, Z. D, Frangi, A.F, and J. yu Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.1, pp.131–137, 2004.

[45] K. S. and R. Y., "Face recognition using 2DLDA algorithm," Proc. of International Symposium on Signal Processing and Its Applications, vol.2, pp.675–678, 2005.

[46] G. Zhao and P. M., "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.6, pp.915–928, 2007.

[47] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual features of intermediate complexity and their use in classification," Nature Neurosci, vol.5, no.7, pp.682–687, 2002.

[48] W. Zhou, A. Ahrary, and S. Kamata, "Image description with 1D local patterns by multi-scans: An application to face recognition," Proc. of IEEE International Conference on Image Processing, pp.4553 – 4556, 2010.

[49] J. Zhang, Y. Wang, Z. Zhang, and C. Xia, "Comparison of wavelet, gabor and curvelet transform for face recognition," Optica Applicata, vol.XLI, no.1, pp.183 – 193, 2011.

[50] T. Mandal and Q.J. Wu, "Face recognition using curvelet based PCA," Proc. of IEEE International Conference on Pattern Recognition, pp.1 – 4, 2008.

[51] H. Huo and E. Song, "Face recognition using curvelet and selective PCA," Proc. of International Conference on Intelligent Control and Information Processing, pp.348 – 351, 2010.

[52] S. Rahman, S. Motahar, A.A. Farooq, and M.M. Islam, "Curvelet texture based face recognition using principal component analysis," Proc. of International Conference on Computer and Information Technology, pp.45 – 50, 2010.

[53] J. Zhang and Y. Wang, "A comparative study of wavelet and curvelet transform for face recognition," Proc. of International Congress on Image and Signal Processing, pp.1718 – 1722, 2010.

[54] Available at http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[55] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, 2007.

[56] P.J. Phillips, P. Flynn2, T. Scruggs3, and K.W. Bowyer, "Overview of the face recognition grand challenge," Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.947–954, 2005.

[57] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," IEEE Trans. Pattern Anal. Mach. Intell., vol.21, no.12, pp.1357–1362, 1999.

[58] D. Xu, S. Yan, and J. Luo, "Face recognition using spatially constrained earth mover's distance," IEEE Trans. on Image Processing, vol.17, no.11, pp.2256 – 2260, 2008.

[59] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG–EBGM," Pattern Recognition Letters, vol.29, no.10, pp.1537–1543, 2008.

[60] A.L.E.G. Manuele Bicego and M. Tistarelli, "On the use of SIFT features for face authentication," Proc. of IEEE International Conference on Computer Vision and Pattern Recognition Workshops, pp.1–7, 2006.

[61] D. Monzo, A. Albiol, A. Albiol, and J.M. Mossi, "A comparative study of facial landmark localization methods for face recognition using hog descriptors," Proc. of IEEE International Conference on Pattern Recognition, pp.1330–1333, 2010.

[62] D. Li, X. Tang, and W. Pedrycz, "Face recognition using decimated redundant discrete wavelet transforms," Machine Vision and Applications, vol.23, no.2, pp.391–401, 2011.

[63] W. Zhou, A. Ahrary, and S. Kamata, "Image description with local patterns: An application to face recognition," IEICE Transactions on Information and Systems, vol.E95-D, no.5, pp.1494–1505, 2012.

[64] W. Zhou, A. Ahrary, and S. Kamata, "Face representation and recognition with local curvelet patterns,," IEICE Transactions on Information and Systems, vol.E95-D, no.12, pp.3078–3087, 2012.

[65] V. Vapnik, "The nature of statistical learning theory," Springer, 1995.

[66] C. Chih-Chung and L. Chih-Jen, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol.2, pp.27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/ c /libsvm.

[67] C. Shan, S. Gong, and P.W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image and Vision Computing, vol.27, no.6, pp.803 – 816, 2009.

[68] R. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Maching Learning, vol.37, no.3, pp.297–336, 1999.

[69] G. Zhang, X. Huang, Y.W. S.Z. Li, and X. Wu, "Boosting local binary pattern (LBP)- based face recognition," Proc. of Chinese Conference on Biometric Recognition, pp.179–186, 2004.

[70] R. Porter and N. Canagarajah, "Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes," Proc. of IEE on Vision, Image and Signal Processing, vol.144, no.3, pp.180–188, 1997.

[71] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.7, pp.971–987, 2002.

[72] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," International Journal of Computer Vision, vol.62, no.1, pp.61–81, 2005.

[73] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.11, pp.2032–2047, 2009.

[74] T. Ahonen and M. Pietikainen, "A framework for analyzing texture descriptors," Threshold, vol.5, no.9, p.1, 2008.

[75] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.8, pp.1265–1278, 2005.

[76] M. Mellor, B.W. Hong, and M. Brady, "Locally rotation, contrast, and scale invariant descriptors for texture analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.1, pp.52–61, 2008.

[77] L. Qin, Q. Zheng, S. Jiang, Q. Huang, and W. Gao, "Unsupervised texture classification: Automatically discover and classify texture patterns," Image and Vision Computing, vol.26, no.5, pp.647–656, 2008.

[78] S. Soottitantawat and S. Auwatanamongkol, "Texture classification using an invariant texture representation and a tree matching kernel," International Journal of Computer Science Issues, vol.8, pp.99–106, 2011.

[79] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," International Journal of Computer Vision, vol.73, no.2, pp.213–238, 2006.

[80] J.C.P. W. Liu, P. P. Pokharel, "Correntropy: Properties and applications in non-gaussian signal processing," IEEE Trans. Signal Process, vol.55, no.11, pp.5286–5298, 2007.

[81] Available at http://www.ux.uis.no/ tranden/brodatz.html.

[82] K.J. Dana, B. Van Ginneken, S.K. Nayar, and J.J. Koenderink, "Reflectance and texture of real-world surfaces," ACM Transactions on Graphics, vol.18, no.1, pp.1–34, 1999.

[83] E. Parzen, "On estimation of a probability density function and mode," The annals of mathematical statistics, vol.33, no.3, pp.1065–1076, 1962.

[84] R. Rockfellar, "Convex analysis," Princeton, NJ: Princeton Univ., 1970.

[85] X. Yuan and B. Hu, "Robust feature extraction via information theoretic learning," Proc. of International Conference on Machine Learning, pp.1193–1200, 2009.

[86] P. Brodatz, "Textures: A photographic album for artists and designers, dover," New York, 1996.

[87] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," Pattern Recognition, vol.43, no.3, pp.706–719, 2009.

Research Achievements

The following lists the main research achievements with published papers.

Jounal

1. Wei ZHOU, Alireza AHRARY, Sei-ichiro KAMATA: Face representation and recognition with local curvelet patterns, IEICE Transactions on Information and Systems, Vol.E95-D, No.12, pp.3078-3087, Dec. 2012.

2. Wei ZHOU, Alireza AHRARY, Sei-ichiro KAMATA: Image description with local patterns: An application to face recognition, IEICE Transactions on Information and Systems, Vol.E95-D, No.5, pp.1494-1505, May 2012.

International Conference Paper

3. Wei ZHOU, Sei-ichiro KAMATA: Linear discriminant analysis with maximum correntropy criterion, Proc. of 11th Asian Conference on Computer Vision (ACCV2012), Korea, Nov. 2012

4. Wei ZHOU, Sei-ichiro KAMATA: Face recognition with learned local curvelet patterns and 2-directional L1 norm based 2DPCA, Proc. of International Workshop on Computer Vision With Local Binary Pattern Variants, in conjunction with 11th Asian Conference on Computer Vision (ACCV2012), Korea, Nov. 2012

5. Wei ZHOU, Alireza AHRARY, Sei-ichiro KAMATA: Image description with 1D local patterns by multi-scans: An application to face recognition, Proc. of 17th

IEEE International Conference on Image Processing (ICIP2010), pp. 4553-4556, Hongkong, Sep. 2010.

6. Wei ZHOU, Alireza AHRARY, Sei-ichiro KAMATA: Face recognition with local feature patterns and histogram spatially constrained earth mover's distance, Proc. of IEEE International Conference on Signal and Image Processing Applications (ICSIP), pp. 374-379, Malaysia, Nov. 2009.

7. Wei ZHOU, Alireza AHRARY, Sei-ichiro KAMATA: Face recognition using local quaternion patterns and weighted spatially constrained earth movers distance, Proc. of the 13th IEEE International Symposium on Consumer Electronics (ISCE2009), pp. 285-289, Kyoto, May 2009.